

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:06:57 ; Search time 5315.66 seconds
(without alignments)
10117.650 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgacgacgaacattacatg.....tctacggttaataactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1848	100.0	1848	6	AX268026	Sequence
2	860.8	46.6	5930	1	AF045245	AX268026 Klebsiell
3	588.6	31.9	10799	1	AE013803	AE013803 Yersinia
4	588.6	31.9	313050	1	AJ414152	AJ414152 Yersinia
5	412.2	22.3	200050	1	AL646068	AL646068 Ralstonia
6	118.2	6.4	10031	1	AF007800	AF007800 Pseudomon
7	113.2	6.1	349116	1	AF003003	AP003003 Mesorhizo
8	113	6.1	3250	8	SCYNR073C	Z71688 S.cerevisia
9	113	6.1	4931	8	SCYNR074C	Z71689 S.cerevisia
10	113	6.1	9172	8	SCS288C14	X86790 S.cerevisia
11	112	6.1	11164	1	AE009374	AE009374 Agrobacte
12	112	6.1	14945	1	AE008240	AE008240 Agrobacte
13	111.4	6.0	50849	8	SCE9669	U18795 Saccharomyc
14	108.4	5.9	10449	1	AE005449	AE005449 Escherich
15	108.4	5.9	26857	1	AP002560	AP002560 Escherich
16	105	5.7	4073	1	AE028937	AE028937 Gluconoba
17	104	5.6	24336	1	AE008841	AE008841 Salmonell
18	103.2	5.6	10446	1	AE000306	AE000306 Escherich
19	103.2	5.6	17093	1	D90849	D90849 E.coli geno
20	103.2	5.6	75888	1	ECOHU47	U00007 47 to 48 ce
21	103.2	5.6	244894	2	AC084324	AC084324 Mus muscu
22	103	5.6	25034	1	AE008844	AE008844 Salmonell
23	103	5.6	230050	1	AL627277	AL627277 Salmonell
24	102.2	5.5	14681	1	AE013893	AE013893 Yersinia
25	102.2	5.5	199050	1	AJ414147	AJ414147 Yersinia
26	101.8	5.5	1458	1	AB063188	AB063188 Gluconoba
27	101.8	5.5	1458	6	AX012443	AX012443 Sequence
28	101.8	5.5	1458	6	E50685	E50685 Gene manipu
29	101	5.5	189370	1	AF010496	AF010496 Rhodobact
30	97.6	5.3	11258	1	AE000251	AE000251 Escherich
31	97.6	5.3	14045	1	D90797	D90797 E.coli geno
32	97.6	5.3	19465	1	D90798	D90798 E.coli geno
33	96.8	5.2	4050	1	EC303143	AJ305143 Brwinia c
34	96	5.2	11783	1	AE005350	AE005350 Escherich
35	96	5.2	286485	1	AP002557	AP002557 Escherich
36	87.6	4.7	9810	1	AF018073	AF018073 Rhodobact
37	84.4	4.6	10955	1	AE005663	AE005663 Escherich
38	84.4	4.6	168764	1	AP002569	AP002569 Escherich
39	84	4.5	4539	1	D13329	D13329 Escherichia
40	84	4.5	11277	1	AE000503	AE000503 Escherich
41	84	4.5	338534	1	ECOUW93	U14003 Escherichia
42	83.4	4.5	10788	1	AE012076	AE012076 Xanthomon
43	82.6	4.5	10977	1	AE004660	AE004660 Pseudomon
44	79.6	4.3	21585	1	AE008765	AE008765 Salmonell
45	79.2	4.3	989	11	CNS06JD2	AL401436 T7 end of

ALIGNMENTS

RESULT 1
AX268026
LOCUS AX268026
DEFINITION Sequence 1 from Patent WO0166779.
ACCESSION AX268026
VERSION AX268026.1 GI:16516559
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Parrott,W., Lafayette,P. and Kane,P.
TITLE Arabitol or ribitol as positive selectable markers
JOURNAL Patent: WO 0166779-A 1 13-SEP-2001;

University of Georgia Research Foundation, Inc. (US)

FEATURES

source 1.1848 /organism="Escherichia coli" /db_xref="taxon:562" BASE COUNT 494 a 443 c 473 g 438 t ORIGIN

Query Match 100.0%; Score 1848; DB 6; Length 1848;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1848; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCCGGCACATCAG 60
DB 1 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCCGGCACATCAG 60
QY 61 GCGTGGTATCTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 120
DB 61 GCGTGGTATCTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 120
QY 121 AATATTGCTAATGATGCTGAACATGCTGATACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 180
DB 121 AATATTGCTAATGATGCTGAACATGCTGATACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 180
QY 181 GTGCTGAAACCGTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 240
DB 181 GTGCTGAAACCGTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 240
QY 241 AAGTTGATACCGTGGCAGGAGATTTACAAACCGCTGATTTGCTGAAGGGGAGATCCGGAAG 300
DB 241 AAGTTGATACCGTGGCAGGAGATTTACAAACCGCTGATTTGCTGAAGGGGAGATCCGGAAG 300
QY 301 ACAAAAGTATGCTTTTACCGTACCGAAGGGGTACTACCTGTAATACCAAGTACACAAA 360
DB 301 ACAAAAGTATGCTTTTACCGTACCGAAGGGGTACTACCTGTAATACCAAGTACACAAA 360
QY 361 CTGGAAGTTAACATCTGATTTAGCGGAGATCTTAAAGGGGAGATCCAAATTTTAC 420
DB 361 CTGGAAGTTAACATCTGATTTAGCGGAGATCTTAAAGGGGAGATCCAAATTTTAC 420
QY 421 GGTGTTATTACCGTATCTCGAAGCGGTATGGCAATTAACCGCGGACCACTAACCCCTG 480
DB 421 GGTGTTATTACCGTATCTCGAAGCGGTATGGCAATTAACCGCGGACCACTAACCCCTG 480
QY 481 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCCGGCACATCAG 540
DB 481 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCCGGCACATCAG 540
QY 541 GCGTGGTATCTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 600
DB 541 GCGTGGTATCTACACCGTTTTCAGAGTGCATGGCGATAAACCGCTGGAGCATTGCTGGCGGC 600
QY 601 AATATTGCTAATGATGCTGAACATGCTGATACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 660
DB 601 AATATTGCTAATGATGCTGAACATGCTGATACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 660
QY 661 GTGCTGGAACCGTACGCGCGAAGGGGTAAAGCAATATGAAGAGATCACTCAATTCAG 720
DB 661 GTGCTGGAACCGTACGCGCGAAGGGGTAAAGCAATATGAAGAGATCACTCAATTCAG 720
QY 721 AAGTTGATACCGTGGCAGGAGATTTACAAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 780
DB 721 AAGTTGATACCGTGGCAGGAGATTTACAAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 780
QY 781 ACAAAAGTATGCTTTTACCGTACCGAAGGGGTACTACCTGAATACCACTACAAA 840
DB 781 ACAAAAGTATGCTTTTACCGTACCGAAGGGGTACTACCTGAATACCACTACAAA 840
QY 841 CTGGAAGTTAACATCTGATTTAGCGGAGATCTTAAAGGGGAGATCCAAATTTTAC 900
DB 841 CTGGAAGTTAACATCTGATTTAGCGGAGATCTTAAAGGGGAGATCCAAATTTTAC 900
QY 901 GGTGTTATTACCGTATCTCGAAGCGGTATGGCAATTAACCGCGGACCACTAACCCCTG 960

RESULT 2

AF045245/c

LOCUS

DEFINITION

AF045245

Klebsiella pneumoniae D-arabinitol transporter (dalT),

D-xylulose-kinase (dalk), D-arabinitol dehydrogenase (dalD), and

repressor (dalR) genes, complete cds.

AF045245 U97126

AF045245.1 GI:2905644

VERSION

KEYWORDS

5930 bp DNA linear BCT 30-AUG-2001

Klebsiella pneumoniae D-arabinitol transporter (dalT),

D-xylulose-kinase (dalk), D-arabinitol dehydrogenase (dalD), and

repressor (dalR) genes, complete cds.

AF045245 U97126

AF045245.1 GI:2905644

VERSION

KEYWORDS

901 GGTGTTATTACCGGTATCTCGAAGCGGTATGCAAAATACCGCGACCACTAACCCCTG 960
961 CTGAATTTGCCATAACGTCGCCCATTAATGGTGAAGTTTCCATGATGCCCTGGTTGAGTTT 1020
961 CTGAATTTGCCATAACGTCGCCCATTAATGGTGAAGTTTCCATGATGCCCTGGTTGAGTTT 1020
1021 CTCAGCTAACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTATGCCCG 1080
1021 CTCAGCTAACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTATGCCCG 1080
1081 AATACCATGTTGACCCCATTAAGCCCTGTCGGCAGCAGAACTTCGGGACAGGATCAAG 1140
1081 AATACCATGTTGACCCCATTAAGCCCTGTCGGCAGCAGAACTTCGGGACAGGATCAAG 1140
1141 GCTCAAAACGGGTATTTGCCGATAAAGCCCGGTAAATGGCGCAAACTTTTATCCAGTGGTC 1200
1141 GCTCAAAACGGGTATTTGCCGATAAAGCCCGGTAAATGGCGCAAACTTTTATCCAGTGGTC 1200
1201 GTGAAGATAAATTTCCGTGATGTCGGCTCCGGCAGTGGAGAGGTGCGTGCAGCACTGGT 1260
1201 GTGAAGATAAATTTCCGTGATGTCGGCTCCGGCAGTGGAGAGGTGCGTGCAGCACTGGT 1260
1261 GCGTCGCTAATCCCTATGAAGAGGCGAAGATTCGCATTTCTTAACCTTTCACACAGTTGC 1320
1261 GCGTCGCTAATCCCTATGAAGAGGCGAAGATTCGCATTTCTTAACCTTTCACACAGTTGC 1320
1321 ATGCGCTGGCAGTACGTTAATCGGTCAAAAATATATCCACGAAAGACCAATGACCGAT 1380
1321 ATGCGCTGGCAGTACGTTAATCGGTCAAAAATATATCCACGAAAGACCAATGACCGAT 1380
1381 TTTATCTATCAGATTTGCCAGCCGCTACGTGACGGAAGATGTCATTTCTTGGTGGCGAT 1440
1381 TTTATCTATCAGATTTGCCAGCCGCTACGTGACGGAAGATGTCATTTCTTGGTGGCGAT 1440
1441 AAGCGTATCATTTGCCAACCTACCGTGATGTTGTACTCAAGCGTTTACCAATCCACAT 1500
1441 AAGCGTATCATTTGCCAACCTACCGTGATGTTGTACTCAAGCGTTTACCAATCCACAT 1500
1501 ATTTCAGGACACCAACCAACCGCTGCGGTGCGGATGTTTCTCGAAATATTCGGCGCATGAT 1560
1501 ATTTCAGGACACCAACCAACCGCTGCGGTGCGGATGTTTCTCGAAATATTCGGCGCATGAT 1560
1561 GCCCCACACTGCGAGAGTGTCTACAGAGGCGGTTCGCCGAGATGCCACCGCATGTTA 1620
1561 GCCCCACACTGCGAGAGTGTCTACAGAGGCGGTTCGCCGAGATGCCACCGCATGTTA 1620
1621 CCGTCACTGTTTACGTTATTCATGGAGCAGTGGCATCACGGCAAACTGCCCTATGAATAT 1680
1621 CCGTCACTGTTTACGTTATTCATGGAGCAGTGGCATCACGGCAAACTGCCCTATGAATAT 1680
1681 CAGGATGGCATCCTTGTATGCACCACTGTCATGCAATGTTTACAGTCTGCCGATCCCGTC 1740
1681 CAGGATGGCATCCTTGTATGCACCACTGTCATGCAATGTTTACAGTCTGCCGATCCCGTC 1740
1741 GCTGTTTATGCCAGTATAAAGCGCTGTTTGGCGATTTAAACCGAAGCTGAAGATTTTGGC 1800
1741 GCTGTTTATGCCAGTATAAAGCGCTGTTTGGCGATTTAAACCGAAGCTGAAGATTTTGGC 1800
1801 GCGTTGTTGGCGGAAAAATCGCTGACGCTACCGGTTAATTAACATAA 1848
1801 GCGTTGTTGGCGGAAAAATCGCTGACGCTACCGGTTAATTAACATAA 1848

SOURCE	Klebsiella pneumoniae.
ORGANISM	Klebsiella pneumoniae
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
	Klebsiella.
REFERENCE	1 (bases 1 to 5930)
AUTHORS	Heuvel, H., Turgut, S., Schmid, K. and Lengeler, J. W.
TITLE	Substrate recognition domains as revealed by active hybrids between the D-arabinitol and ribitol transporters from Klebsiella pneumoniae
JOURNAL	J. Bacteriol. 179 (19), 6014-6019 (1997)
MEDLINE	97464425
PUBMED	9324246
REFERENCE	2 (bases 1 to 5930)
AUTHORS	Heuvel, H., Shakeri-Garakani, A., Turgut, S. and Lengeler, J. W.
TITLE	Genes for D-arabinitol and ribitol catabolism from Klebsiella pneumoniae
JOURNAL	Microbiology 144 (Pt 6), 1631-1639 (1998)
MEDLINE	98304087
PUBMED	9639934
REFERENCE	3 (bases 1 to 5930)
AUTHORS	Heuvel, H. and Turgut, S.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-1998) Biology, University of Osnabrueck,
	Barbarastr. 11, Osnabrueck, NS 49076, Germany
COMMENT	On Feb 23, 1998 this sequence version replaced gi:2735580.
FEATURES	Location/Qualifiers
source	1..5930
	/organism="Klebsiella pneumoniae"
	/strain="1033-5P14"
	/db_xref="taxon:573"
	/clone="KAY2026"
gene	complement(421..1698)
	/gene="dalr"
CDS	complement(421..1698)
	/gene="dalr"
	/codon_start=1
	/transl_table=11
	/product="D-arabinitol transporter"
	/protein_id="AAC26500.1"
	/db_xref="gi:2905647"
	/translation="MSINNKQMLPLNLWGYIAIAVFMGTGDFGLAFSLHYIKALG ESPAESAPFLYLGLAALSAMISGVVAELITPLKTMIGVLVHCYVHLELVFLGLH ANYALILFYGRGAPFLFLYSEIIVAIHVNKSDNASSAIGWFVAIVTSIGVFGSY IPSTIPIHIGMGFTLWALAFCLTGGVIALVSLRHQTPQHMLNLTREKFESELRGA TLLYTNRIILSSMVRILNTSLFQFAVIMPEVDELGFSTSEWLQVWVFFFTTIF SNVLGILGEKLGKVMKVVWFGICGMAISLAFYIIPHFCHSPAMALIPIAALGIFV AAFPVLAAPALPDKHGAISVYNLSAGNSNLPALPAIAVLLPFFSTIGVVIATYA LYVAFELCAPIRVEQPFQFSKHEATREQVEFS"
gene	complement(1769..3232)
	/gene="dalr"
CDS	complement(1769..3232)
	/gene="dalr"
	/codon_start=1
	/transl_table=11
	/product="D-xylulose-kinase"
	/protein_id="AAC26499.1"
	/db_xref="gi:2905647"
	/translation="MYLIGDLTSEVRALVDENHEVIAHSAPLSIORPHPHVSEQA PELWATEYLMATLREKCAQHPAIRAIGLSGOMHGVALLDAGEKALRPAILWNTDR CAESCALEANPELVQAGNLMPGTPAPKLLVRRHPQHFORTATVLLPKDLYRY RMTGKSDMSDAAGTLWLDVAKRDSMDALDKCLRSQMPKLVGECEVSLDPOV AARGLNASVNVAGGGDNVSAIGVAVSPGDAFI SLGTSGLVLFVVDAYRPAQSA VHAFCHVLPNLWHQMSVMSAICLQWFCRLTGTTEVALLAEIAELSEDKANAPFL PYLSGERTPHNDPARGIFWGMTHASLRAQLGYAVLEGVSFGINDGLQALKESGPIA QCSLVGGGARSFPWQLLADILAMPVVTTHKGGTGGALGAARLACLAAGKPIAAVCEK PEVQWTRADPIRHHTLMRYAQKALYLNDLHVRQH"
gene	complement(3242..4609)
	/gene="dalr"
CDS	complement(3242..4609)
	/gene="dalr"
	/codon_start=1
	/transl_table=11
	/product="D-arabinitol dehydrogenase"

protein_id="AAC26498.1"	
db_xref="gi:2905646"	
translation="MNOFTWLHIGLGSFHRHAQWALHRLIASGDNWRITAAAGNIRN DAEQVQALAAQGGRYVLETVSPGEEYEEITSTQKLLPQAGLOPLINGANPQTK LNCNVRHNGRFDHGMVEFLQTLQKQAVIDMAATPCNTMDRTIPRAADLPAR IKQAQIDDKAPVAGETFIQWVNFNFRPNLEAVGVMEVSPYEEAKIRILNA SHKSIWAGTLIGQYIHTSFTIYVIAADRYVTVDICLGDNGIDLPTIRDVVLK RFTNPYIQDTNORVAADGFSKIPAMIAPTLQECYQGVRRPEATAMLPAFLFVEQW KGTLPYQYQDGLIDQAQVHMFQAQDPVAVFARDKALFGDLANNADFLALMRKVA YTLIN"	
4816..5757	gene
/gene="dalr"	
4816..5757	CDS
/gene="dalr"	
/codon_start=1	
/transl_table=11	
/product="repressor"	
protein_id="AAC26497.1"	
db_xref="gi:2905645"	
translation="MSKEDDIRLDQKVRAAWMYIAGQNSQSEIASOLGTSRPVVORLI AAKEGIVSNLHPVANCIDYQQLQEKYGLTECNVVPAPFSESTLDSVSFGCYOL MARYLODDKEKILGSLGLTKALORIDFSLATRCVALLISAMNADQOCNYDDVPL LLTRKIRAKYIOWPAPRYAQADEFEMCTNRLRFSVSGVAREADVIFVIGLPGTQK PIFKDGFINQOMDELTAARGIGELGRFIDAGDGVVDSEINRMITSYDIHQSHCPRI AACGGEHKRPAILAALKGGWINGLVTDHTARWLLTR"	
1303 a 1789 c 1720 g 1118 t	BASE COUNT
ORIGIN	
Query Match	46.6%; Score 860.8; DB 1; Length 5930;
Best Local Similarity	76.8%; Pred. NO. 6.2e-259;
Matches 1051; Conservative	0; Mismatches 317; Indels 0; Gaps 0;
Qy 481 ATGAACGAACAATTTACATGGCTGCACATCGGGTAGGTCTTTTCATCGGCACATCAG 540	
Db 4609 ATGAACAATCAATTCATCGGCTTCATATCGGTCTGGGTCTTTTCATCGGCACATCAG 4550	
Qy 541 GCGTGTATCTACACCGTTTGCAGGTGATGGCGCATAAACGCTGGAGCATCTCTCGGGC 600	
Db 4549 GCGTGTATCTCATCTGCTGATTCTTCGGGGGATAATCGCTGGCGCATCGCGGGG 4490	
Qy 601 ATATTCGTATGATGCTGAACATGCTGTACAGGCACCTCAGTCACACAGAAAGTCCCTAT 560	
Db 4489 AATATTCGCAACGCCGAGCAGGTGGTCCAGGCGCTGGCGCGCAGGAGGACGTTCAC 4430	
Qy 661 GTGCTGGAACCGCTCAGCCCGGAAGGGGTAACGGAATATGAAGAGATCACCTCAATTCAG 720	
Db 4429 GTCTCGAGACCGTCAGCCCGGAAGGGGAGCGGAATATGAAGAGATCACCTCAATTCAG 4370	
Qy 721 AAGTTGATACCGTGGCAGCAGATTTTACAACCGCTGATTGTGTAAGGGGAGATCGCGAAG 780	
Db 4369 AAATGCTACCGTGGCAGCGCGGTACAGCGCTGATTAAAGAGGGGCAAAACCGCGAG 4310	
Qy 781 ACAAAAGTGATTGCTTTCACCGTCACCGAAGCGGGGTACTACTGTAATACCAGTCACAA 840	
Db 4309 ACCAAAGTGATTGCTTTCACCGTCACCGAAGCGGGGTACTACTGTAATACCAGTCAC 4250	
Qy 841 CTGGAAGTTAAACAATCTTGATTAGCGGCAGATCTTAAAGGGGGATGCAAAACAATTTAC 900	
Db 4249 CTGGAACACCAACCCCGATCTGAGCGCGACCTGCAAGGCGAGTGCACAAACCATTTAC 4190	
Qy 901 GGTGTTATTACCCGTATCTCGAAGCGGTATGGCAAAATAACGCCGAGCACCTAACCCCTG 960	
Db 4189 GGCACCCCTCGCGCGGATCTCTGAAAAGCGCATGGCGGACAAACCGCGCGCTGACCCCTG 4130	
Qy 961 CTGAATTCGGATAAGTCGGCCCATATGTGAACGTTTCCATGATGCGCTGTTGAGTTT 1020	
Db 4129 CTCAACTCGGATAAGTCGGCCCATATGCGCGAGCGCTTTCACGACGCTATGTCGAGTTC 4070	
Qy 1021 CTCACGACTAACGGCAACACAGGATGTCAATCGACTGCTGAGTACAATAACCTTCGCCG 1080	
Db 4069 CTGACGACTCACCGCAACACAGCGGGTGATTGACTGGATGGCGCCCAATACCACCTGTCG 4010	

QY	1081	AATACCATGTTTACCGCATTTACCGCTCTGTCGGCAGCAGAACTTCGGGCACCGATCAAG	1140
Db	4009	AACACCATGTGACCGCATTCACCCCGCTGTCGGCGCCGATGTCGGCCCGCATCAA	3950
QY	1141	GCTCAACGGGTATTCGCGATAAGCCGCGTAAATGGCGAAACCTTTATCCAGTGGGTC	1200
Db	3949	GCCGACCGGATTTGATGACAAAGCCGCGTATGGGGAGACCTTTATCCAGTGGGTA	3890
QY	1201	GTGGAAGATAATTTCCGTGATGTCGTCGGCAGCTGGAGAGGTGCTGTCGAACCTGGTG	1260
Db	3889	GTGGAAGAACTTCCCGCATGTCGCCCGCAATCTCGAGGACGTCGCGGTGGAGATGGTG	3830
QY	1261	CGTCGGTAATCCCTATGAGAGGAGAGATTCGCTTAACTCTTCACACAGTTGC	1320
Db	3829	GAGTCGGTATCCCGTATGAAGAGGGAATCCGTATTTTGAACGCTCGCACAGCTGC	3770
QY	1321	ATCCCTGGCAGGTACGTTAATCGTCAAAATATATCCACGAAAGCAATGACCGCAT	1380
Db	3769	ATTGCTTGGCGGAACCTTAATCGCCAGCAGTATATCATGAAAGCAGCTGACCGAT	3710
QY	1381	TTTATCATCATGATTCCGACCGCTACGTGACGGAAGATGTCTATCTCTTGGGCGAT	1440
Db	3709	GTTATCTACGCCATTCCGACCGCTACGTTACGAGGAGCTTATTCGCTCTCGGCGAC	3650
QY	1441	AACGGTATCGATTGGCAACCTACCGTGATGTTGTACTCAAGCGTTTACCATCCACAT	1500
Db	3649	AACGGCATTTGATCTCGGACCTACCGGGATGTGTGCTTAAAGCGCTTACCAACCCCTAT	3590
QY	1501	ATTGAGACACCAACACGCGTCGTCGGGATGTTTCTCGAAATTCGCGGATGATT	1560
Db	3589	ATCAGAGACCAACACGCGTCGCGCGGCGGCTTCTCGAAATTCGCGGATGATC	3530
QY	1561	GCCCCACATCGGAGAGTCTACACGAGGCGTTTCGCCGAATGCCACCGCATGTTA	1620
Db	3529	GCGCCAACCCTGACGAGTGCTACACGCGCGGCTCCGCCGAGCGCGCATGCTG	3470
QY	1621	CCTGCACTGTTTACGATTTATGAGAGCAGTGCCATCAGGCAAACTGCCATGATAT	1680
Db	3469	CGGCGCTGTTCTCTCTTTATGAGCAGTGCGCAACGGAACCTCTGCCATATCAGTAC	3410
QY	1681	CAGGATGGCATCTTGATGCACCATGTCCTCATGCAATGTTACAGTCTGCCGATCCGTC	1740
Db	3409	CAGGAGGCATCTCGATGCGAGCGGTCGATGAGATGTTTCGAGCTCAGGATCCGGTG	3350
QY	1741	CGTGTATTGCCAGTGATAAAGCGCTGTTTGGCGATTAAACCGAAGTGAAGATTTGCC	1800
Db	3349	GCTGTCTCGCGCGTGATAAAGCGCTGTTTGGTGATCTGCCAACAACGCGATTTCCTG	3290
QY	1801	GCCTCTGTCGGCGAAATACGCTGACGCTACGCGCTTAATTAACATA	1848
Db	3289	GCCTGATGCGGAGAAAGTCGCCCGCGTTTACACGCTGATTAACTAA	3242
RESULT 3			
AE013803/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
1 (bases 1 to 10799)			
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,			
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,			
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,			
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,			
Blattner, F. R. and Perry, R. D.			
Genome Sequence of <i>Yersinia pestis</i> KIM			
J. Bacteriol. 184 (16), 4601-4611 (2002)			
12142430			

REFERENCE		2 (bases 1 to 10799)	
AUTHORS		Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,	
		Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,	
		Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,	
		Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,	
		Blattner, F. R. and Perry, R. D.	
TITLE		Direct Submission	
JOURNAL		Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445	
		Henry Mall, Madison, WI 53706, USA	
FEATURES		Location/Qualifiers	
source		1..10799	
		/organism="Yersinia pestis KIM"	
		/strain="KIM"	
gene		/db_xref="taxon:187410"	
		267..605	
		/gene="emrE"	
		/note="y2000"	
CDS		267..605	
		/gene="emrE"	
		/function="transport; drug/analog sensitivity"	
		/note="residues 3 to 112 of 112 are 50.90 pct identical to	
		residues 1 to 110 of 110 from E. coli K12 : B0543;	
		residues 3 to 112 of 112 are 54.54 pct identical to	
		residues 1 to 110 of 110 from GenPept : >gb AAC64463.1	
		(U67194) QacE [Enterobacter aerogenes]"	
		/codon_start=1	
		/transl_table=11	
		/product="methylviologen resistance"	
		/protein_id="AA085566.1"	
		/db_xref="GI:21958828"	
		/translation="MIMSFVYLFMAITAEVATTMLKASDGFSLRVPVVVIGVI	
		APWGLSQVVKTMPLGIAYIWSGLGVLVIAATFTMYQKLDMAAVVGMALIIISVMV	
		INLLSKTPMH"	
		complement(665..1324)	
gene		/gene="y2001"	
		complement(665..1324)	
CDS		/gene="y2001"	
		/function="putative transport"	
		/note="residues 1 to 195 of 219 are 50.25 pct identical to	
		residues 1 to 195 of 215 from E. coli K12 : B3508"	
		/codon_start=1	
		/transl_table=11	
		/product="putative transport ATPase"	
		/protein_id="AA085567.1"	
		/db_xref="GI:21958829"	
		/translation="MITDLRLITLAGILGLERQLRAKEAGLIRTHILVIGSAM	
		FMVSKYGFEDLLVLEHVSFDPSSVAOVVSGMFLGAGTIQKQVKGTLTAAGMW	
		VTAAGLVIGSLYEIGIYGLTMTLVLEVFRLQSRLGLGHQYIVLVKLLPESVPLVL	
		VALQKLHLPNLSVTQNHKSDNTSELTMEVTLSPKTVNQVYEQLLAIEIHQLDI	
		K"	
		1537..1950	
gene		/gene="y2002"	
		1537..1950	
CDS		/gene="y2002"	
		/function="unknown"	
		/note="residues 87 to 134 of 137 are 56.25 pct identical	
		to residues 1 to 48 of 51 from GenPept : >dbj BAAL4988.1	
		(D90776) ORF_ID:02646; similar to [SwissProt Accession	
		Number P52646] [Escherichia coli]"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical"	
		/protein_id="AA085568.1"	
		/db_xref="GI:21958830"	
		/translation="NMVYIILAFINRYLISLSCRKMRIFQGVVCGMALFLAACSS	
		NDSNTLNNSNTLANSSTYEPQQATNAIAYRNVMNSNDVNSAANCANAGAIL	
		ANTRQLHGSVGMCPDQKRCDEALLRGACPV"	
		complement(2008..2460)	
gene		/gene="hslJ"	
		/note="y2003"	
		complement(2008..2460)	
CDS		/gene="hslJ"	
		/function="factor; adaptations, atypical conditions"	

[illegible]

QY	859	GATTTAGCGCAGATCTTAAAGGGGATGCAAAACAATTTACGGTGTATTATCCCGTATC	918
Db	15310	GATATTACGGCCGATCTAGAGAGATTGCCGACGATTTTACGGGCAATTACCGCTATC	15369
QY	919	CTCAAGCGCGTATGCAAAATACCGCGACCACTAACCTGCTGAATTCGGAATAACGGT	978
Db	15370	TTCAACAACGGCAACAGGATAGATGGCCCGGTCACTCTACTGAACCTCGGATACTTG	15429
QY	979	CGCCATAAGTGAAGCTTTCCATGATGGCTGGTTGAGTTTCTCCAGCTAACTGCAAA	1038
Db	15430	CGTCACAACGGTCAACGATTTCCGCCATGGCTTATTAGAGTTTCTCTCTACGAGTCAA	15489
QY	1039	CAGGATGTATCAGTCAGTGGCTGAATACAAATACACATTTGCCGAATACCATGGTTGACCG	1098
Db	15490	CAAAGCCTATTAAACTTGGGTGACAAGCCAGACCCGAGCCGAATACCATGGTTGATCGT	15549
QY	1099	ATTACGCTCGTCCGCGCAGCAGAACTTCGCGCAGCATCAAGCTCAAAAGGGTATTGCC	1158
Db	15550	ATCACCCGCGCCCATCAGCGCGATATCGCGCAGCGGTGTGGAANAACCGGTATCGTA	15609
QY	1159	GATAAAGCGCGGTAAATGGCGCAACCTTTATCCAGTGGGTCTGGAAGATAATTTCCCGT	1218
Db	15610	GATAAGCCTCGGTGATGGGGAATCATTTATTCAATGGGTATTGAGGATGATTTTATT	15669
QY	1219	GATGTCGTCGCGCACTGGAGAGTCGCTGTCGAACCTGGTGGCGTCGGTAATCCCGTAT	1278
Db	15670	GCTGGCGCGCGCATGGAGAACGTTGGGCGGAATGGTCACTCGGTTTGGCCCTAT	15729
QY	1279	GAAGAGCGGAAGATTCCGATTTCTTAACCTCTTACACAGTTGCACTCGCTGGCGAGTACG	1338
Db	15730	GAAGAAGCAAAATCCGCATCTGAACGCCAGCCACAGTTGATTGCTGGGCGGCAGC	15789
QY	1339	TTAATCGGTCAAAAATATATATCCAGGAAGACAAATGACCGATTTTATCTATCAGATGGC	1398
Db	15790	TTGATTGGGCAATGTTATATTATCAAGAGTACCAACCGCAGCTATCCGCCAGATGGCT	15849
QY	1399	GACCGCTACGTGACGAAGATGTCATTCTTCTTGGCGGATACAGGTATCGATTTGCCA	1458
Db	15850	TATGACTATGTCACGAAGATGTCATTCTCCCTCCCTGACGCCACCGACACTTGATTAGCC	15909
QY	1459	ACCTACCGTGATGTTGACTCAAGCGTTTTTACCAATCCACATATTCAGGACACCAACCA	1518
Db	15910	CATTACCGGATGTTGACTGGAACGTTTCAGTAACTCGGTATATTCGCGATACGAACCG	15969
QY	1519	CGCGTCGTCGCGATGGTTCTCGAAATTCGCGGATGATTCGCCCGACACATGCGAGAG	1578
Db	15970	CGTCTCGCCCGCATGGATTTTCCAAATCCCGGATTTATCACTCCCGACGCTGATCGAG	16029
QY	1579	TGCTACAGCGGCGGTTCGCGGAATGCAACCGCATGTTACCTCGACTGTTTTCAGTA	1638
Db	16030	TGTTACCAAGTGGCGAGTGCTACCGCTACCGCAATGTTGCCCGCATTTATCTTCGTA	16089
QY	1639	TTCAATGAGCAGTGGCATCACGCAACCTGCCCTATCAATATCAGGATGCATCTTGAT	1698
Db	16090	TTTATGAAAGTTGGCAACCAAGCACTCTGCCGTATGATATCAGATGATTTCTGGAC	16149
QY	1699	GCACAGCTGTCCATGCAATGTTACAGTTCGCGATTCGCGCTGCTTTATGCGCACTGAT	1758
Db	16150	CCAGCAGCGCTCCATGCCATGTTTCACTCTCTGACCTTCAGGCTTATTGCCCCGTGAT	16209
QY	1759	AAAGCGCTGTTTGGCGATTTAACCGAAGCTGAAGATTTTGGCCCGTGTTCGCCGAAAA	1818
Db	16210	AGCGCCCTTTTCGGGCGGTGGCCAAATGATGATCAGTTCGAACAGTATTATACCCAGTCC	16269
QY	1819	ATCGCTGAGCTCTACCGGTTAATTAAC	1845
Db	16270	ATAAATCGTCTCTCTCTGGGTAGC	16296

RESULT 5
AL646068/c 200050 bp DNA linear BCT 07-DEC-2001
LOCUS Ralstonia solanacearum GM1000 chromosome, complete sequence:
DEFINITION

segment 12/19.
AL646068 AL646052
AL646068.1 GI:17429060
Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 200050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 200050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CERPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Garbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers.
1. .200050
/organism="Ralstonia solanacearum"
/strain="GM1000"
/db_xref="taxon:305"
complement(85. .573)
/gene="RSC2042"
/note="RS03607"
complement(85. .573)
/gene="RSC2042"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD15749.1"
/db_xref="GI:17429061"
/translation="MRVIESLDALRGLIQEVAVSDWVEITQQOVNQFEATGDRDWI
HYDVERARRESFPGAVAHGFLTLLPALMQSALDMPDKMGVNTGLNKRFPAPVP
YSLRLARVSLLEMEMLPLQSPDAPALTAQMTWGTVECEGQARPCVPAESISR
Ys"
complement(600. .1070)
/gene="RSC2043"
/note="RS03608"
complement(600. .1070)
/gene="RSC2043"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"

QY 735 GCAGGAGATTTACACCGCTGATTTCTGAAGGGGAGATCCGACACAAAGTATGTC 794
Db 94273 GTCCGCGATCTTGGCGCGCTCATCAATACGGGCGGATCCGCGTGCAGTCTGTC 94214
QY 795 TTTACCGCTCACGAAGCGGTACTACCTGAATACCACTACAACTGGAAGTTAACAA 854
Db 94213 GTTTACGGTACCAAGGCGGTATTACCTTCGACGACGATCGTCTCGATGTCACGA 94154
QY 855 TCTGATTTAGGGGAGATCTTTAAAGGGGATGCAAAATTTACGGTGTATTACCCG 914
Db 94153 TCCCGATCTCGGCGCGACCTTCGGGCGCGCTCGACGCTATACGGCGATTCGCCCG 94094
QY 915 TATCTCGAAGCGGTATGCGAAATTAACCGCGGACCACTAACCCCTGCTGAATTCGCA 974
Db 94093 GTTGCTCGCGAGCGCGGACGCGGAGCGCGGCGCTTCACGCTCAGAGTTGCACAA 94034
QY 975 CGTGCCCATATATGTTGAACGTTTCCATGATGCGCTGTGAGTTCTCCAGCTACTCG 1034
Db 94033 TCTGCGCAGCAGCGCGCGGTTTTCGCGGGGATGCGCGGTTCTCGCGTGCAGCG 93974
QY 1035 CAAACAGGATGTCATGACTGCTGAGTACAAATACCACTTCCCGAATACCATGTTGA 1094
Db 93973 CGACGAGCGCTCTCGCGTGTTCGATGCCAACGTTATCGTCCGCGGCGATGTCGA 93914
QY 1095 CCGCATTAGCCCTGCTCCGCGCAGCAGAACTTCGGGACGATCAAGCTCAACGCGTAT 1154
Db 93913 CCGGATTCACGCGCGCGCGCAGCAGCAGCTCCGCGCGCTGCACGCGGCCACCGCGT 93854
QY 1155 TGCAGTAAGCGCGCGTAAATGGCGAAACCTTTATTCAGTGGTCTGCGAAGATAATT 1214
Db 93853 CGACGACCGCTCGCGCGTATGCGCGAATCTTTCATCCAAATGGTGATCGAGACAAT 93794
QY 1215 CCGTGATGTCGCTCCGCGCAGTGGAGAGTTCGTTGCGAATGGTGGCGTCGTAATCCC 1274
Db 93793 CATCGCGCGCGCGCATGGAGATCGCGGCGGAGATCGTCCGCGATGTGATCC 93734
QY 1275 CTATGAAGAGGAGATTCGATCTTAACTCTTCACACAGTTGATCGCTGGCGCAGG 1334
Db 93733 GTACGAAGGAGGAGATCCGATCTGTAATGCGAGCGCACAGTGTGATCGCTGGCGG 93674
QY 1335 TAGCTTAATCGGTCAAAAATATATCCAGAAACAGCAATGACCGATTATATCATCAGAT 1394
Db 93673 CAGCTCGCGGACTCACCTACATTCACGAAGGATGCGTGATCGCGCATTTATCGCT 93614
QY 1395 TCCGACCGCTACGTGACGGAAGATGTCATCTTCTGCTTGGCGATAACGGTATCGATT 1454
Db 93613 CGCCTACGACTACGTACCGACGACGTGATTCCTCGCTGACGCGGCGCGTGGATCT 93554
QY 1455 GCCAACCTACCGTATGTTGTAAGCGCTTTTACCAATCCACATATTCAGGACACCAA 1514
Db 93553 CGAGCGCTATCGGGAGCTGCTGCTCGAGCGCTTCGSCAATCGTACGTCTCGACACCA 93494
QY 1515 CCAACCGCTCGCTGCGGATGTTTTCGAAAATTCGCGGATGATTGCCCGCACACTCGG 1574
Db 93493 CGAGCGTTCGCGCGCGCGGATTCCTCGAAAATCCCGGCTTCATCGCACCGCGCTCG 93434
QY 1575 AGATGCTACCAAGGAGGCTTGCCTGCGGAAATCCACCGCATGTTACCTGCACTGTTTA 1634
Db 93433 CGAATGCTTGGCGCGCGCGGATTCCTGCGGACCGCGGTGTTGCGCGCGCTGTTCT 93374
QY 1635 CGTATTCATGAGCAGTGCATCAGGCAAACTGCGCTATGAATATCAGGATGGCATCT 1694
Db 93373 GGGCTTCCTTGAAGTGTGGCGCGCGGACCTTGCCCTACGTATACACGAGGCGGTGAT 93314
QY 1695 TGATGCACCACTGTCATGCAATGTTACAGTCTGCGGATCCGCTGCTGTTTATGCCAG 1754
Db 93313 GATGCGCGCGCGCAGCAGCATGCTCGAGGCTCGGATTCGTTGCGCGCTTCGTTC 93254
QY 1755 TGATAAGCGCTGTTTGGCGATTTAAACCGAAGTGGAAGATTTTGGCGCGCTTGTTCGCGCA 1814
Db 93253 AGATGCCCAATTTGTGGGTTTCGCTTGGCGCGCGCAGCGCTTGTCCAGCGGTGCGTGC 93194
QY 1815 AAAATCGCTGAGCTACGCGT 1837

Db 93193 GGGACGCGCGCGGTTGAAGCAT 93171

RESULT 6
AF007800
LOCUS
DEFINITION
Pseudomonas fluorescens mannitol operon, mtlE (mtlE), mtlF (mtlF),
MtlG (mtlG), MtlK (mtlK), mannitol dehydrogenase (mtlD), xylulose
kinase (mtlY), and fructokinase (mtlZ) genes, complete cds.
AF007800.1 GI:2293413
VERSION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
Pseudomonas fluorescens.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 10031)
Brunker,P., Altenbuchner,J., Kulbe,K.D. and Mattes,R.
Cloning, nucleotide sequence and expression of a mannitol
dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
Escherichia coli
Biochim. Biophys. Acta 1351 (1-2), 157-167 (1997)
97236441
9116029
JOURNAL
MEDLINE
PUBMED
REFERENCE
2 (bases 1 to 10031)
Brunker,P., Altenbuchner,J. and Mattes,R.
Structure and function of the genes involved in mannitol, arabinol
and glucitol utilization from Pseudomonas fluorescens DSM50106
Gene 206 (1), 117-126 (1998)
98121321
9461423
JOURNAL
MEDLINE
PUBMED
REFERENCE
3 (bases 1 to 10031)
Brunker,P., Altenbuchner,J. and Mattes,R.
Direct Submission
Submitted (10-JUN-1997) Institute of Industrial Genetics,
University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany
On Aug 4, 1997 this sequence version replaced gi:2065484.
LOCATION/Qualifiers
1. .10031
/organism="Pseudomonas fluorescens"
/strain="DSM50106"
/db_xref="taxon:294"
814. .842
932. .2242
/gene="mtlE"
932. .2242
/gene="mtlF"
/note="similar to E. coli periplasmic maltose binding
protein Malt; periplasmic subunit of the periplasmic
binding protein dependent mannitol transport system"
/codon_start=1
/transl_table=11
/product="mtlE"
/protein_id="AAC04468.1"
/db_xref="GI:2293414"
TPEASHPEIKLNVLEENLVLRQRTLDIATOGQDFVLTIGMVEAALMGAKWLEPM
KDLPSYDLDDVFPSPVRSVGLSVKGLYALPFYAESSITVYRTDLEKAGLSMPEHPTW
SQIGFAAKLTDKSKEOYGLCLRGKAGCNMALJTLANGYARWDEKQWPEFNGP
EMKDALNFYVDMKSKSGPPGSSNGFNENLALFNSGKCAIWDASVAGSFVDTKQSK
VADHVGFAPHERTDKGTSLWLSLAIPSTSSKAKDAKVTSWATSKYEGALVAKT
DGVAWPPGTGRKSTYSDEYMAAPKATVLESKLVADPTKTLKVPVYIGIQLVTYPE
FOAIGTVGKFPFGALTGQTVDAALTAQTTEREMKRAGYPK"
2381. .3271
/gene="mtlF"
2381. .3271
/gene="mtlF"
/note="similar to E. coli MaltF protein; inner membrane
subunit of the periplasmic binding protein dependent
mannitol transport system"
/codon_start=1
/transl_table=11

promoter
gene
CDS

gene
CDS

QY	1033	GGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTTGC	CGCCGAATACCATGGTT	1092
Db	5885	CACAACCCGAGCTGCATGACTGGATCAAGGCCCATGTGAGCTTCC	CGGAACGCATGGTC	5944
QY	1093	GACCGATTACGGCTCGTCCGGCAGCAGAACTTCCGGCAGCGATCA	AGGCTCAACCGGTT	1152
Db	5945	GACCGCATACGGCGATGACACAGCACCGCCACCGCTGCAACTGC	ACGATGAACACGG	6004
QY	1153	ATTGCCATAAAGCGCGGGTAATGGGGAACCTTTATCCAGTGGGT	CGTGGGAAGATAAT	1212
Db	6005	ATGACGATCCCTGGCCAGGTGTTTGGGAACCCCTTTGTGCAGTGG	GTTACTGGAAGACAAA	6064
QY	1213	TTCCGATGATCGTCCGCGCACTGGGAAGTGGGTGTCGAACCTGT	GCGTGGCGTAATC	1272
Db	6065	TTGCTCAACGCGCCCGCGGTGGGAAAGGTTGGCGTGCAAGTTC	ACCGACGATGTGACA	6124
QY	1273	CCCTATGAAGAGCGGAAGATTCGCATCTTAACTCTTACACAGTGC	ATTCAGCTCGCGTGGGCA	1332
Db	6125	CCCTATGAAGAGATGAAGATCGGCTTGCTCAAGCGGACGCCACCT	GGCCCTGACCTACCTG	6184
QY	1333	GGTACGTTAATCGTCAAAAATATATCCAGAAAGACATGACCGATT	TTATCTATCAG	1392
Db	6185	GGTTTCTCAAGGCTATCGGTTGTGCACGAGACCATGAACCGCTGT	TCGTGGGCC	6244
QY	1393	ATTGCCACCGCTACGTGACGGAAGATGTCATTCTCTTGGT---	GGCGGATAACGGTATC	1449
Db	6245	TACATGCGCGCTACATGAGCCTCGAGCTCAGCCAAACCTCGCGCG	GTTACCGGGCATC	6304
QY	1450	GATTTGCCAACACCTACCGTGATGTTGTTACTCAAGCGTTTACC	AAATCCATATTCAGGAC	1509
Db	6305	GACCTGACCGACTACAGCAGACAGCCCTGGTGGACCGCTCTCTC	CAACACGAGCGATTGCCGAC	6364
QY	1510	ACCAACCAACCGCTCGTGGGATGTTCTCGAAATTC	CGCGGATGATGCCCCAC	1568
Db	6365	CAGTTGAACGGGTGTTCCGATGGCTCGTCAAGTTTCCCAAGTTC	ACCGTCCGAC	6423
RESULT 7				
AP003003				
Mesorhizobium loti DNA, complete genome, section 10/21.				
AP003003.2 GI:14024067				
Mesorhizobium loti (strain:MAFF303099) DNA.				
Mesorhizobium loti				
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
Phyllobacteriaceae; Mesorhizobium.				
1 (sites)				
Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,				
Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,				
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,				
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpou,S., Sugimoto,M.,				
Takeuchi,C., Tamada,M. and Tabata,S.				
Complete genome structure of the nitrogen-fixing symbiotic				
bacterium Mesorhizobium loti				
DNA Res. 7 (6), 331-338 (2000)				
JOURNAL				
MEDLINE				
PUBMED				
2 (bases 1 to 349116)				
Kaneko,T.				
Direct Submission				
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research				
Institute, The First Laboratory for Plant Gene Research; Yana				
1532-3, Kisarazu, Chiba 292-0812, Japan				
(E-mail:kaneko@kazusa.or.jp,				
URL:http://www.kazusa.or.jp/rhizobase/,				
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)				
On May 11, 2001 this sequence version replaced gi:11994978.				
COMMENT				
FEATURES				
source				
1..349116				
/organism="Mesorhizobium loti"				

/strain="MAFF303099"				
/db_xref="taxon:381"				
complement(54..1655)				
/gene="ml13875"				
complement(54..1655)				
/gene="ml13875"				
/codon_start=1				
/transl_table=11				
/product="phosphoglycerate dehydrogenase"				
/protein_id="BAB50672.1"				
/db_xref="GI:14024068"				
/translation="MAPRVLVSDKLSPTAVQIFKORGVEVDYLDLGGKDEKLEVID				
QYDGLAIRSATKTEKLIINAATRLKVIAGRGVDNDIPAAASRGKGIYVNPFGNSI				
TTAEHAVAMIFALAQIPENASTHAGKWEKRMFGVEITGKTGLVIGCGNIGSIVAT				
RGVGLKMHVAFDPFLSDKRAEELGVEKVELDELFAADFITLHTPLTKTRNIIDAG				
ATAKMGVRIINCARGGLIVEADLIAALKSGKAGIDVEVEPAEONALFGMENV				
VATPHLGASTAEQENVAQAEQMSDYLIKAVNSAINWPSITAEERPLKPFVKLA				
EVLGAFVQVTEDEIMEVELFDGSTATMNTALISATLAIAGLAIQVSDVNMVSAPI				
VKEGIIIAEVRKDSGVDFGIKLTVKTEHRTSIAGTCFSDGKPRFIQIKGINLDA				
EYQGHLYTTNADAPGIIIGLGTGCGGVNIANFOLGRNRPGGDAIALLYLDAPPE				
KYLEQVRAHKSIDSARKLQFDVGGI"				
complement(1719..2894)				
/gene="ml13876"				
complement(1719..2894)				
/gene="ml13876"				
/codon_start=1				
/transl_table=11				
/product="phosphoserine aminotransferase"				
/protein_id="BAB50673.1"				
/db_xref="GI:14024069"				
/translation="MTHTKPLGRPANPNFSSGPCAKRFGWSVEALKNAALGRSHRAK				
IGKTLQEAIELTREILQVPADHRIGIVPASDTGAVEMALWSLLGERGVDMAWESFG				
SGWTVDVVKQLKPDVKLEAGYGLDPLAKIDFDRDVTWNGTTSVGVVPNGDFIP				
ADKGLTICDATSAFAORLDFEKLDVVTFSQKVLGGGAGHMLTILSPRAVARESY				
KPAWLPKIFRLTSGKLEIGFKETINTPMLCVEDYLDLHAKSGTGGDLVLAR				
ADANAALDGFVDKSWGLAVPATRSNTSVCLSTFDLPDAALDADGQATFAKGLV				
SALDKEGVAYDIGSYRADPPGLRIWCGATVETSDLETLLPDLWDAFAAQAKSLKAA"				
/gene="ml13877"				
complement(3096..3869)				
/gene="ml13877"				
/note="hypothetical protein"				
/codon_start=1				
/transl_table=11				
/protein_id="BAB50674.1"				
/db_xref="GI:14024070"				
/translation="MTLSRIVLALAGLLPLTPAVAADYDPPIVADQAPVVPVEV				
GSGWYLRGDVSYLAQSKFNDDPFTPPASDFEKEDEPIFASIGGYHFNDRYLRADNLG				
YLPGNKIGIGYDLSVVPPTATSVASADLKNATYSLMUNATYDLGTYYGTTPYLGCG				
YGVOSTRRLSANFYTNNGDPTDDEVDQDDTKYSLATYTLNAGLAYQVSKNYSVDLGY				
QYFSAPDAEYVTAASLTSPVHKGISNHQIKLGLRYDLW"				
complement(4111..4944)				
/gene="ml13878"				
complement(4111..4944)				
/gene="ml13878"				
/note="hypothetical protein"				
/codon_start=1				
/transl_table=11				
/protein_id="BAB50675.1"				
/db_xref="GI:14024071"				
/translation="MFNTARKALLALLFAGLASPVFAADLPQVEEAPPPVEQPQVD				
VGGWYIRGDIYHKSVDYIDMTYTGIDPCNCTVPGGKSFYGLKGLGSLGGVGY				
KINDRLRADVADYWFKSNFNGTSDLVSTEVSKMSALLANAYVIGTWGHTTP				
YVAGIGGARVKWDTYDPNTAETNPNGSNRFAVALMAGASYCLTDKTLIDAGYRFS				
HYGGGRMFEDWASSAGPGDRGINTHEVRGGLRYQFGGNGNCAAPVVAQPEPEPIYT				
K"				
complement(5193..6545)				
/gene="ml13879"				
complement(5193..6545)				
/gene="ml13879"				
/codon_start=1				
/transl_table=11				
/product="phosphoglucomutase/phosphomannomutase family				

protein: MrsA"
/db_xref="GI:14024072"
/translation="MAGNYFGTGDIGIRANKFPMTAIAIRVGMAGLSFQGRSHRH
VVLGKRLSCYMTENAMVAGLCAAGMDVLLGPITPAVAMLVSRADIGVMSAS
HNPYDNGIKLFGPDGYKLSDEIEERIESMLDKDVELTLADSDGLGRAKRVGDVHDY
IEFARKTIRMSLSLGURI VVDCANGAAYKVAPEALWELGAEVVAIINIEPNFNIKE
CGSTHAGLQKHVEVRADIGIALDGDADRVVIYDENGAIVDGQIMAMIAESWHQS
RLAGGVYVMSLGLERLGMKQLQHRTKVGDYVVEHMRHGLNMGVEQSGHIV
LSDFSTTGDGLVALQFLACIKROGRVPSLSKFFEPVOLLKNVRITAGGKPLEEAPV
KAATEDARNRLGRAGRLVIRPSGTEPLIRVMAEGDDPQLVEAVNDIVEISETRSAA
"
6913. 7779
/gene="mlr3880"
6913. 7779
/gene="mlr3880"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50677.1"
/db_xref="GI:14024073"
/translation="MLIHRLATLGLAVVTLFLAAPANALTMACESTKYNAAKDAGTL
AGOTWNOFRACQCTDAATTTDKPATTKAAETPKARATETKPAKAAAADSTAKGL
TSKESAKYQAAKAGTLNGMKWDFRTECAACASAAUTTKPATTKAAAP
AGGGLTMAESAKYQAAKANTLQKQWDFRKAEGCAAADSDVPASEANTYE
PAKPTTVAPKGVTPPSAISPKYASETEPKGRMHTCLDOYYALKDANALGLKWTQKG
GFYSLCNAKLKG"
8079. 8537
/gene="mlr3881"
8079. 8537
/gene="mlr3881"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50678.1"
/db_xref="GI:14024074"
/translation="MPNLKLAVALTASVAVSAPASALTMKECEGYRYRAKEGGTLN
GMDAEPKKEKCATPAAESVSGDVTTPQPKETDAVAPATVAPAGVPTFTLLAEF
KWTAPAKRMKTCQLQGYHGNKEAGTLNLRWIKQGGGYISLCNARLKAES"
complement(8624. .10552)
/gene="mlr3882"
complement(8624. .10552)
/gene="mlr3882"
/codon_start=1
/transl_table=11
/product="metalloprotease (cell division protein); FtsH"
/protein_id="BAB50679.1"
/db_xref="GI:14024075"
/translation="MNNRYNMLAWIAIVLLIALFNLFQTPQTRGASSDVPYSQFLQ
DVAAGRVKTVIAGRTIVTDNSTGFQYSPGDQLVSLRLODKNVTINARPEADGS
NSLFGYLSLPMILILGWIFFMRQMSGSRAMGFKSKAKLILTEAHGRVTFQDVA
GVDAKEDLEIVELFQPKQFORLGGKIPRGVLLVPGGTGKTLARSAVAGANVP
FTISGSDVEMFVCGASRVDMFDOAKNAPCIIFIDEIDAVGHRGAGLGGNDER
EQTNLQLLVENDGESNESIILIAATNRPDVLDPALLRPRFRQVVPNDIYGRER
ILKVHVRNPLAPNLDLVKVARGTGFGSADLMNLVNESALMAARNRKLRTVMAEFD
AKKIMGAERSAMTQAEKELTAYHEAGHAILALNVPSSADPLHKATITPRGRALGH
VMQLPEDGRSMYSKYMSRLAIMMGVRVAPEFNGENITSGASSDIEQATKLARM
VTRWGFSDKLGHVAGDNOEVEFLGHSVARTONISEETAQIIDEAEVRLIDEAYSTAK
SILTKKKKWLALAQGLLEYETLSGDEIKOLIAEKPARDLGDPTPSRGSAVPSKSG
RRKXGPEPEGMPEQPS"
complement(10675. .12039)
/gene="mlr3884"
complement(10675. .12039)
/gene="mlr3884"
/note="Similar to YaeN protein [Bradyrhizobium japonicum],
contains similarity to cell cycle protein MesJ/cytosine
deaminase-related protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50680.1"
/db_xref="GI:14024076"
/translation="MLDTEPDLSTRLFSHIDFTHGAVAAVSGSDSTALLLLHHFED
RTGSSAKLLAVTIDHGLRQGSAAEAQAVAKLCAERGIHAHRLTLTWRKRPSTGLPAAAR

DARYHLLAEARABEGILIVTGHHTADDOAETVLMRHRDRELADLAGRLAGMAPATL
YDREWIVRPLLTGRRSALRDFLREHVHGWAEADPTNIDEAFPRVRAALAGSAGHM
ENTURLAQAAVERGQLGASALIRRIASQSTGLIRLDPALLIADQAAIYALRIL
LATAGGAFLPDQASAAFLVRLKAGLALTSRTVVDFRHAGFLRREARGLPPAAT
AVDNTINDGRRHITLNDMSGALLTAPLCAAAARLADDDGETPASLIRAAALAEPTLL
QAFENLGLPQGEPLRPVFAARPVVSPFAFLPSFDLAPARAVAEILGNASPLPASPLGG
HSAD"
complement(12098. .13210)
/gene="mll3886"
complement(12098. .13210)
/gene="mll3886"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50681.1"
/db_xref="GI:14024077"
/translation="MHLRSVLSGTLAALLSFTALASGPGASGPGSTDSGFSFHLPL
SLELPSLFGDKKPDQVQFAQOSNGAGATGLEDQLROMNGKIEELNFOILQOMRMRK
AQEDNEFRFOLEGGAGGPPKPKSDATNTNTDVAAPATQAPADAGAAPGNQS
OGKTEVEDVIVSESGDPKVPITGTAPEKTFGSITVDKNGNIDAGANTQSTAPQD
AAPATKAPAKAGKNGTIVAAPLSTNDPEELYRNSYQFILSGDYGTAQGPRDRHSRF
"
Query Match 6.1%; Score 113.2; DB 1; Length 349116;
Best Local Similarity 46.9%; Pred. No. 3.5e-23;
Matches 524; Conservative 0; Mismatches 558; Indels 36; Gaps 4;
Qy 503 TGCACATCGGGTTAGTCTTTTCATCGCGCACATCAGCGGTGTATCTACACCGTTTCG 562
Db 132260 TGCATCTCGCGCTCGCGCCITTCATCGCGCCCATCAGCGCCCTATGTGCGACGAGTGCC 132319
Qy 563 AGGTGATGGGGATAAACGCTGGAGCATTGCTGCGGGCAATATTCGTAATGATGCTGAAC 622
Db 132320 TGGCGGGGGGAAACGAGCTGGGCATACAGGGGTCTCTCTTGGCGACCGCCGACACGC 132379
Qy 623 ATGTGCTACAGCACTAGTGCACAGAAAGGTCTGTCTGTCTGGAACCGTCGCCCGG 682
Db 132380 GCGATCGCGTGTGCG-----CGCAGAGCGGCTCTACACGTGGGATCCCGACGAGTG 132433
Qy 683 AAGGGTAAAGCAATATGAAGAGATCACCTCAATTCAGAAGTTGATACCGTGGCAGGACG 742
Db 132434 GCGAGGA---GGAGCTGTGCTCATCGGCTCTATCCAGTCGATGCTGCTGGCGCGGAAA 132490
Qy 743 ATTTACAACCGCTGATTGCTGAAGGGGCGAGATCCGAAGACAAAAGATGTTGCTTTCACCG 802
Db 132491 ACCCCAGCGCGTGTGCGCGCTGACCGATCCACGACCGCCGATCGTCACGCTGACCA 132550
Qy 803 TCACCGAAGCGGGTACTACCTGAATACAGTCACAACTGGAAGTTAAACAATCCTGATT 862
Db 132551 TCACCGAAGCGCTTATCTCAGGGCTGCGGGCGCGGATTTGGACGCGCCCATCCGACA 132610
Qy 863 TAGCGGCAGATCT-----TAAAGGGGGATGCAAAACAATTTACGCTGTTATTACCCGTA 916
Db 132611 TCATCCAGATCTGCCCAATCCGCAATCCCGAGGCGCGCATGGTTCTTCGCGGAGG 132670
Qy 917 TCCTCGAAGCGGTATGGCAATAACGCGGACCACTAACCCTGCTGAATTCGGATAACG 976
Db 132671 CGCTTTCGCGCAGACTGGCGCGCGCACCCGCCCTTACCGCTGTGCTGCGACAACC 132730
Qy 977 TCGGCCATAATGGTGAACGTTTCCATGATGGCTGGTTCAGTTTCTCCAGCTA----- 1029
Db 132731 TCGCGGCCAACGGCGCCACGCTGCATCGGCTGCTCCACCAATTCGCAACGCGCGATG 132790
Qy 1030 -----ACTGGCAACAGGATGTCTACGACTGGCTGAGTACAAATACCAATACCAC 1075
Db 132791 CGAGTCTACCACTGGAGCGACACAGAAATCCCGACCATATCGCGAACACGAGTCCGCT 132850
Qy 1076 GCCCGAATACCATGTTGACCGCATTTACCGCTCTCGCGGACGAGAACTTCGCGGACGGA 1135
Db 132851 TTCCTCGAGCATGTCATCGCATCGTGC CGGCAACCAACCGACCGCGCGCGCGCA 132910
Qy 1136 TCAAGGCTCAACAGGGTATTTGCCGATAAAGCGCGGTAAATGGCGGAAACCTTTATCCAGT 1195
Db 132911 TCGCTGGGAACTCGGTGTGACGATGCTTGGCCTGTCTATGATCCGAGCCCTTCCGCAAT 132970

```
QY 1196 GGTGCTGGAGATAATTTCCGTCATGTCCTCGGCACCTGGAGAGGTGCGGTGTCGAAC 1255
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132971 GGTGCTGAAGACAATTTCTCCCGCGCAGCGGACTGGGAAGATTGCGGCTACCA 133030

QY 1256 TGTGCGGTGCGTAATTCCTCATGATGAGAGCGGAGAGTTCGCATCTTAACCTTTCACACA 1315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133031 TGTGTCGCGATGTCGGCCCTTCGAGGACATGAAGCTCAGGCTGCTCAAGCGGCGCAT 133090

QY 1316 GTTGATCGCCTGGCGAGGTACGTTAATCGTCAAAAATATATCCAGAAAGCACATGA 1375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133091 CGGCTACGCTATCTCGGCTGCTCAGCGCCACCGCTCGACCGCGCCTTTGCGG 133150

QY 1376 CCGATTTCATCATGATTCGCGACCGCTACGTGACGGAAGATGTCATTCCTTGTG 1435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133151 ATCCGGCGATCCGCAATTTGTCAGCGGCTCTGGCGGACGATCCCGACATTCGCG 133210

QY 1436 GCATACGCTATGATTCGCAACCTACCGTGATGTTGCTACGAGGCTTTTACCAATC 1495
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133211 AGGATCGGCGATGATACGCGCAACCTACACAGCGAGCTTGCAGCGCTTTCTCAACA 133270

QY 1496 CACATATTCAGGACACCAACCAACGGCTCGCTCGGATGTTTCTCGAAATTCGCGCGA 1555
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133271 CCGGCTTGGCCACCGCAGCGGCGATGCCAATGACGCGACCGCAACTCCGCGAGC 133330

QY 1556 TGATGCGCCACACTGCGGAGTGTACACGAGGCGC 1593
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133331 GCATCATCGGCTCGGCGATCGAGTGCTCGAAGCGCGC 133368

RESULT 8
SCYNR073C/c
LOCUS SCYNR073C 3250 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNR073c.
ACCESSION 271688 Y13139
VERSION 271688.1 GI:1302609
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 3250)
Andre,B., Iraqui Houssaini,I., Urrestarazu,L.A. and Vissers,S.
Unpublished
REFERENCE 2 (bases 634 to 3250)
Levesque,H., Lepingle,A., Nicaud,J.M. and Gaillardin,C.
Unpublished
REFERENCE 3 (bases 1 to 3250)
MIPS.
Direct Submission
TITLE Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
source Location/Qualifiers
1. .3250
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XIV"
/complement(441..1949)
/note="ORF YNR073c"
/codon_start=1
/protein_id="CAA96356.1"
/db_xref="GI:1302610"
/db_xref="SWISS-PROT:P39941"
/translation="MTKSDTTATSLNAKTLKSFESTLPIPTYREGVKQGVHLGVG
AFRSHLAVFMRHLKQWHLKWSICGVLKMDALMDKAAQSEALNLTSLVGRINDT
NAYTVGSITAYAFDDPRAVTERMANPDTHIVSLVTENGYSHSCATSLMTDAPEI
INDLNHEPKFDLYLYEALLRYKRLPTIMSCNDMPQGVTVKTMVFAFLK
KDEKFAEMEDRVTPNSMVDRTVPRCTDKERYVADTWGINKDQCPVVAEPFIQWLE
DNFSDGRPPWELGVQVVKDVSLELMKRLNLLGGHSAMGYLGLAGVYIHEVNDP
TINKYIRVLMREVIPLLPKVPGVDFEYTSVLFERSNPATQDVARICLMSGKMP
KVLPSIYDLRKPDKYKLLAVCVAGHFRILYTGVDMNGKPEIEDHPAPTLLKAAVK
```

```
GGKDPHELLNIEVLFSPEIRDNKEFVAQLTHLSLETVDYDKPIAAIKEILDQV"
BASE COUNT 968 a 651 c 638 g 993 t
ORIGIN
Query Match 6.1%; Score 113; DB 8; Length 3250;
Best Local Similarity 49.8%; Pred. No. 1.6e-23;
Matches 315; Conservative 0; Mismatches 315; Indels 3; Gaps 1;
QY 949 CCACTAACCTGCTGTAATTCGATAACGTCGCAATAATGGTCAACGTTTCCCATGATGC 1008
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1379 CCATTCACATATATGATGTCGCAACATGCCCAAAATGGTGTACAGTAAGACCATG 1320

QY 1009 CTGTTTTCAGTTTCTCCAGCTAACTGGCAACAGAGTGTATCGACTGGGTGAGTCAAAAT 1068
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1319 CTGTTTTCAGTTTCTCCAGTTTAAAGAAAGGATGACAAATTCGCCGCTGGATTGAACAAG 1260

QY 1069 ACCATTCGCCGAATACCATGTTGACCGCATTTACGCTGTCGCGGAGCAGAACTTCG 1128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1259 GTTACTTCTCTAACAGCATGTTGGACCGTGTGACCCACGTTGTACCGGATAAGAGCGT 1200

QY 1129 GCACGGATCAAGGCTCAACACGGGTATTGCGGATAAAGCGCGGTAAATGGCGCAACCTTT 1188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1199 AATACGTTGCTGACACCTGGGGAATCAAGGATCATGTCCTGTCGAGAACCTTTC 1140

QY 1189 ATCCAGTGGGTGCGGAAGATAATTTCCGTGATGTCGTCGCGCACTGGAGAGGTCGT 1248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1139 ATCCAATGGTCTTGAAGACAACATTTCTCGATGGCGCTCTCCATGGGAACCTTGTG 1080

QY 1249 GTCGAATGTCGCTCGGTAAATCCCTATGAAGAGCGGGAAGATTGCACTTCTTAACCT 1308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1079 GTTCAGGTGCTCAAGGATGTCGATTCCTACGAATTTGATGAAGTTGCGTCTACTTAA 1020

QY 1309 TCACAGATTGTCATGCCCTGGGCGAGTACGTTAATCGGTCAAAAATATATCCACGAAAGC 1368
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1019 GGACATTCCTGCTATGGATATTTGGGATACCTTGGCAGGCTACACCTATATACATGAGTT 960

QY 1369 ACAATGACCGATTTTATCATGATTCAGATTGGCGACCGCTAGCTGACGGAAGATGCTAT 1428
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 959 GTCAACGACCCAACTATCAACAAGTATATCCGTTGTTGATGCGTGAGGAAGTTATCCCA 900

QY 1429 TGCTTCG---GGCGATAACGCTATCGATTGTCGAACCTACCGTGATGTTGCTACAGCGT 1485
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 899 TTATTCGCTTAAAGTGGCAGGTGTTGATTTGGAAGAGTACACTGCATCAGTGTGGAAAGA 840

QY 1486 TTTACCAATCCACATATTCAGGACACCAACCAACGCTCGCTCGGATGTTTCTCGAAA 1545
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 839 TTTCTCAATCCAGCTATTCAGGACACTGTGCGACGATTTGTTGATGGGCTCTGCTGAAG 780

QY 1546 ATTCCGGCGATGATTGCCCGCCACACTGGGAGAG 1578
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 779 ATGCTTAAGTATGTTTGGCCATCAATTTAGGAG 747

RESULT 9
SCYNR074C/c
LOCUS SCYNR074C 4931 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNR074c.
ACCESSION 271689 Y13139
VERSION 271689.1 GI:1302611
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 4931)
Andre,B., Iraqui Houssaini,I., Urrestarazu,L.A. and Vissers,S.
Unpublished
REFERENCE 2 (bases 1 to 4931)
Levesque,H., Lepingle,A., Nicaud,J.M. and Gaillardin,C.
Unpublished
REFERENCE 3 (bases 1 to 4931)
MIPS.
Direct Submission
TITLE Submitted
```

Db 326 GTCAACGACCAACTATCAACAAGTATATCATCGTGTTTGGTGTGAGGAGTATATCCCA 267

QY 1429 TGCTTG---GGCGATAACGGTATCGATTGTTGCCAACCTACCGCTGATGTTGTACTCAAGCGT 1485

Db 266 TTATTGCCTAAAGTGCAGGTGTGATTTCGAAGAGTACACTGCATCAGTCTTGGAAGA 207

QY 1486 TTTACCAATCCATATATTCAGGACCAACCAACGCGTCGCTGCGGATGGTTTCTCGAAA 1545

Db 206 TTCTCCAATCCAGCTATTCAGGACACACTGTGCACAGTATTGTTGTTGATGGGCTCGTAA 147

QY 1546 ATTCCGGCGATGATTGCCCCACACACTCGGAG 1578

Db 146 ATGCTTAAGTATGTTTGGCCATCAATTACGAG 114

RESULT 10

SCS288C14/c

LOCUS SCS288C14 9172 bp DNA linear PLN 27-NOV-1995

DEFINITION S.cerevisiae N3810, N3815, N3820 and N3825 genes.

ACCESSION X86790

VERSION X86790.1 GI:805055

KEYWORDS S288C.

SOURCE Saccharomyces cerevisiae.

ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 9172)

AUTHORS Levesque,H., Nicaud,J.M., Lepingle,A. and Gaillardin,C.

TITLE Sequence of a 9.2kb telomeric fragment from the right arm of S. cerevisiae chromosome XIV

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 9172)

AUTHORS Nicaud,J.M.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-1995) J.M. Nicaud, Institut National de la Recherche Agron., Genetique Moleculaire et Cellulaire, Thiverval Grignon, F- 78850 Thiverval Grignon, FRANCE

FEATURES

source 1..9172

/organism="Saccharomyces cerevisiae"

/strain="S288C"

/db_xref="taxon:4932"

/chromosome="XIV"

/map="right telomere"

/clone="pEL186B"

/complement(1..1316)

/gene="N3810"

/complement(<1..1316)

/gene="N3810"

/codon_start=1

/protein_id="CAA60486.1"

/db_xref="GI:805056"

/db_xref="SWISS-PROT:P39941"

/translation="MTKSDETTATSLNAKTLKSFESTLPIPTPREGVKOGIVHLGVGAFYHSLVFMHRLMDEHHLKSGVGLMKADALMRDAMKAQDCLYTLVERGIKDTNAYIVGISTAYMAPDPRAVIEKMANPDTHIVSLTVTNGYVHSEATNSLMTDAPELINDLHPEKPDLYLYEALLRYKRLTPTTMSCDNMPQCVTKTLVAFAKLKDFEKAFTEDKVTSPNSMVDVYTPRCTDKERYVADTWGCKDQCPVVAEPFIOWVLEDNFSDGRPHWELGVQVKVDVDSYELMKLLRLLNGHSHAMGYLGTYAGTYTIVHVNDDPKINKYIRLVMREVIPLLPKVPQDFEETASVLERFSNPAIDTVAIRICLMGSGMKPKYVLSPYIEQLRPDGKIKLLAVCVAGWFRYLTVGDMNGKPFTEIDP"

gene

CDS

complement(2618..3754)

/gene="N3815"

complement(2618..3754)

/gene="N3815"

/codon_start=1

/protein_id="CAA60487.1"

/db_xref="GI:805057"

/db_xref="SWISS-PROT:P52923"

/translation="MTINTKNIYVVGAGVGVSVANHLRYELGGTYAIKLVYASNYVFLPSAVLTVSKDQTYLILPKNLVDSGIEVIKDTAASFDDKEVLGSDRAIKFDILVLAATGSKWADIGSTYTGDNKYFEFEARSISADHILFGLGGFVNCLEAGELLFYSLEIRSGKRISITIHNSDKLLPDSGLYNDTLRKNVTDYLSKNGITLYLNTVGLASLDTSL

gene	PKRIELGESSKYIDADLIYRGVISPVPVNSISDLCDKKGFIQVEKFRKAVEAG NVFAIGDVTNRYHGLVRDNMVDVLTRNVSSIQEGTEASLVADADCLTGHAPSGVS LGNAGFQGFPLPLLTGINIPISFLISRAKSNLFSDKMEPLFKK"											
	/gene="N3820" 4938..6062											
	CDS	/gene="N3820" 4938..6062 /codon_start=1 /codon_start=1 /protein_id="CAA60488.1" /db_xref="GI:805058" /db_xref="SWISS-PROT:P52924" /translation="MGEEIRLLSSKGLTHSSFCPLPEDIFRNRLTWLCSERNSGFWI WSLSWPLAVWGMSTWYPLLTSLVILIALIILPVQIQSYKTLTSLNQLTQUSKE LIKSPAGSADMDVAIFHSYLVENKAWKTAHFFNGTDCQEAFTKTLLEPAVLRR QNEDRFSEIILPVYTEDAVQYFTKVAQNRHLHGKCKEKLMLNWKLPKETYRC KLAWSCQIRISFDLPDMSDFICANLSQACPLVSVLWIDIFFRMIDDFQNIIRPK SMKVDKQWYLSDIINEOGASPEKWDITLARKNTRLFERRVMKNEBEFFDGTDCQAF ERNFSSILFSKKSASPRSLNVLWKYIQEAQLSCNVEPLP"										
		gene	6940..7302 /gene="N3825"									
			CDS	6940..7302 /gene="N3825" /codon_start=1 /codon_start=1 /protein_id="CAA60489.1" /db_xref="GI:805059" /db_xref="SWISS-PROT:P52921" /translation="MYKLTISAAGVAAIAATASATTLAQSDERNVLVELGVVVSIDIR AHLAQYMFQAAHPETYPVEVAEAVFNVDFTTTLTGTPADQVTRMITGVPWYSTRL KPAISKALSKDGIYITIAN"								
BASE COUNT	2794 a	1728 c		1908 g	2742 t							
ORIGIN												
Query Match 6.1%; Score 113; DB:8; Length 9172; Best Local Similarity 49.8%; Pred. No. 2e-23; Matches 315; Conservative 0; Mismatches 315; Indels 3; Gaps 1;												
QY	949	CCACTAACCCCTGCTGAATTCGCGATAACGTCGCCCAATAATGGTCAACGTTTCCATGATGCG	1008									
Db	746	CAATTCACATATTATGTCATGTGACACATGCCCAATAATGGTGTACAGTAAAGACCATG	687									
QY	1009	CTGTTGAGTTTCTCCAGCTAACTGGCAACAGGATGTCATGACGTGGCTGAGTACAAAT	1068									
Db	686	CTTGTTGCATTTGCCAAGTTAAAGAGGATGAGAAATTCGCCCGCTGGATTGAAGACAAG	627									
QY	1069	ACCATTGCGCGAATACCATGGTTGACCGGATACGCCCTGCTCGGCGAGCAGAACTTCCG	1128									
Db	626	GTTACTTCTCTAACACATGCTGGACCGTGTGACCCACGCTGTGTACCGATAAAGAGCGT	567									
QY	1129	GCACGGATCAAGGCTCAACAGCGTATTGCCGATAAAGCGCGTAAATGGCGCAACCTTT	1188									
Db	566	AAATACGTTGCTGACACCTGGGGAATCAAGGATCAATGTCCTGTCGACGAACCTTTC	507									
QY	1189	ATCCAGTGGGTGCTGGAAGATAATTTCCGTGATGTCGCTCCGCGACTGGAGAAGTCCGT	1248									
Db	506	ATCCATGGGTCTTGAACACACTTCTCCGATGGCGCTCCCTCCATGGCAACTTGTGCT	447									
QY	1249	GTCGAACCTGGTGGCGTCGGTAATCCCTATGAAGAGCGCAAGATTCGCATCTTCTTAACCT	1308									
Db	446	GTTTCAGGTGCTCAAGGATGTCGATTCCTACGAATTCATGAAGTTGCGTCTACTTAACTG	387									
QY	1309	TCACACAGTTGCATCGCTGGGCGAGTACGTTAATCGGTCAAAAATATATCCACGAAGC	1368									
Db	386	GGACATTTCTGCTATGGGATATTTGGGATACCTTGGCAGGCTACACCTATATACATGAGTT	327									
QY	1369	ACAATGACCGATTTTATCATCAGATTCGCGACCGCTACGTCGAGGAAGATGTCATCTCT	1428									
Db	326	GTCACGACCCCACTATCAACAGTATATCCGTGTTTTCGATGGTGAGGAAGTTATCCCA	267									
QY	1429	TGCTTG---GGCGATAACGGTATCGATTTGCCAACCTACCGTGATGTTGTTACTCAAGCGT	1485									
Db	266	TTATTGCTTAAAGTGCCAGGTGTTGATTTTCGAAGAGTACACTGCATCAGTGTGTGGAAGA	207									
QY	1486	TTTACCAATCCACATATTCAGGACACCACCAACGCGTCGCTCGGATGGTTCTCTCGAAA	1545									

Db	206	TTTCTCCAAATCCAGCTATTTCAGGACACTGTCGACGTAATTTGTTGATGGCTCTGGTAAG	147
QY	1546	ATTCCGGCGATGATTCGCCCGCCACACTGCGAGAG	1578
Db	146	ATGCCTAAGTATGTTTGGCCATCAATTTACGAG	114
RESULT 11			
AE009374			
LOCUS			
DEFINITION	Agrobacterium tumefaciens str. C58 (U. Washington) linear BCT 20-DEC-2001 of 187 of the complete sequence.		
ACCESSION	AE009374 AE008689		
VERSION	AE009374.1 GI:17742928		
SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington).		
ORGANISM	Agrobacterium tumefaciens str. C58 (U. Washington)		
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
AUTHORS	1 (bases 1 to 11164)		
	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.		
TITLE	The genome of the natural genetic engineer Agrobacterium tumefaciens C58		
JOURNAL	Science 294 (5550), 2317-2323 (2001)		
MEDLINE	21608550		
PUBMED	11743193		
REFERENCE	2 (bases 1 to 11164)		
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, J.Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA		
FEATURES	Location/Qualifiers		
source	1. .11164		
	/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"		
	/strain="C58"		
	/db_xref="taxon:180835"		
gene	94..1578		
	/gene="mtlK"		
	/note="Atu4451"		
CDS	94..1578		
	/gene="mtlK"		
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"		
	/codon_start=1		
	/transl_table=11		
	/product="mannitol 2-dehydrogenase"		
	/protein_id="AA15245.1"		
	/db_xref="GI:17742928"		
	/translation="WTCKLISLATLDEAKKTAAPAYSRNDLSAGIVHFGVGNFRAHQ AVYLDLFTNGADHPALIGACVLPDSATMBREKLAQDFLTTVSQDDNNATCARVTGP MDILKVGDTQALIDTLADPKTRIVSMVTITEGGYFIDASGSFNPOHPATADECKNPSA PDKTVFIVAGLRKNDQLQDPTWNSCDNIHPNGKTVKNAVVGAAUSDPAFANNWII		

QY	802	GTACCGAAGCGGGTACTACCTGAATACAGATCAACAACCTGGAAGTTAACAAATCCTGAT	861
Db	484	ATTACCGAGCGGGTTATTTCATCGATCGCTCGCGTCTCTTCAATCCGACGACATCGGCC	543
QY	862	TTAGCGGCAGATCTTAAAGGGGATGC-----AAACAATTTACGGTGTATTATACCCGT	915
Db	544	ATTGCCGAAGCGCGCAGAACCCGCTCGCTCCGAAACCGGTTCGGTTCATCGTGTGCC	603
QY	916	ATCCTCGAAGCGGGTATGGCAATAACCGCGACCACTAACCCCTGCTGTAATTCGGATAAC	975
Db	604	GGGCTGAAGCGCGCGCGATAGGACCTGACGCCCTTCACCGTCATGTCCTCGACAAAC	663
QY	976	GTCCGGCATTAATCGTGAACGTTTCCATGATGGCCTCGGTGAGTTTCTCCAGCTAACTGGC	1035
Db	664	ATTCCCAACAATGGCAAGGTGACGAAACCGCGGTGTCGGGCTGCGAGCGTCTC---C	720
QY	1036	AAACAGATGTCATCGACTGGCTGAGTACAAATACCACTTGCCCGCAATACCATGGTTGAC	1095
Db	721	GACCCGCTTTCCCAACTGATCGCGGAAATGTGCTTCCGAAATTCATGCTGTCGAC	780
QY	1096	CGCATTACCGCTCGTCCGGCAGCAGAACTTCCGGCACGATCAAGGCTCAAAACGGGTATT	1155
Db	781	CGCATCACCCGCCACCGCGAGCGGTGAACGCAATATCGCCGTCGACGATTTCCGGCATC	840
QY	1156	GCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCAGTGGTCTGCGAAGATAATTTC	1215
Db	841	GAGGATAATTTGGCGGTCTTCTCGAAGAGTTTCAAGCAATGGGTGATGGAGGATAATTTC	900
QY	1216	CGTGATGTCGTCGGCAGCTGGAAGGTCGGTGTGCGAACTGGTGGCGTGAATCCGCC	1275
Db	901	CCGCGAGCGCTCCGGGCTCGAAGAAGCGGGCTCCAGTTCGTCGAAGATGTTGGCGCC	960
QY	1276	TATGAAGACGCGAAGTTCGATCTTTAACTCTTCACACAGTTGCATCGCCTGGGCAGGT	1335
Db	961	TACGAGCACAATGAAGATCCGCATCTCTGAATGGTGGCCACGCGCAATCGCCTATCCGCCA	1020
QY	1336	ACGTTAATCGGTCAAAATATATCCACGAAGACCAATGACCGATTTTATCTATCATGATT	1395
Db	1021	CGCTCTCGACATTCATTTCGTGTCATGAGCGGATGAGCATCCCGTGTATCCGGCGGTTT	1080
QY	1396	GCGGACGCTACGTACGGAAGATGTCATTCTCTGTT---GGCGGATAACGGTATCGAT	1452
Db	1081	CTGCGCAAGCTCAAAAGACGAGATCTCCCGGTCTATTCGCGCGGTACTTGACACCGCA	1140
QY	1453	TTGCCAACCTACCGTGTGTTACTCAAGCGTTTACCAATCCACATATTCAGGACACC	1512
Db	1141	CTCAACGACTATTTCGGGCTGATCGAGCGCGCTTCTCTCAACCCGAAGATTGCGGATACC	1200
QY	1513	AACCAACGGCTCGCTCGGATGTTTCTCGAAATTCGCGCGATGATGCCCCCACCACATG	1572
Db	1201	ATTCCGCGCTCGCGAGGATGGTTCACACCGCCGAAATTCATCTCTGCGCTCCACT	1260
QY	1573	CGAGAGTGTACACGAGCGGCG	1594
Db	1261	TTAGATCGGCTGTCGCGCGCGC	1282
RESULT	12		
AE008240/c			
LOCUS	AE008240	14945 bp	DNA linear BCT 18-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 44 of 187 of the complete sequence.		
ACCESSION	AE008240	AF007870	
VERSION	AE008240.1	GI:15158760	
KEYWORDS			
SOURCE	Agrobacterium tumefaciens str. C58 (Cereon).		
ORGANISM	Agrobacterium tumefaciens str. C58 (Cereon)		
	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
REFERENCE	1 (bases 1 to 14945)		
AUTHORS	Hinkle, G., Slater, S.C. and Goodner, B.		
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall		

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

Disease in Plants

Unpublished

2 (bases 1 to 14945)

Hinkle, G., Slater, S.C. and Goodner, B.

Direct Submission

Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA

Approximately 800 bp of telomeric sequence missing from the left end of the chromosome and 200 bp missing from the right end.

Location/Qualifiers

1..14945

/organism="Agrobacterium tumefaciens str. C58 (Cereon)"

/strain="C58"

/db_xref="taxon:181661"

98..865

/gene="AGR_L_810"

98..865

/note="HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN GLPC-AIS INTERGENIC REGION"

/codon_start=1

/transl_table=11

/product="AGR_L_810p"

/protein_id="AAK88977.1"

/db_xref="GI:15158761"

/translation="MTDSDSERYRAPALDKGLDLELLELARTDGGTLQIEIAKAIGKSP
NELYRLDLVRGVYQDRDGRFSLTLKMEGLAHFHAPIRRLVFAAPVMDRFSASA
EQACHLAVYDRGSVVVIAQESATYAMSMRVGAQMSFHTSGHVLAFQTEAOREI
MTEQVRGSGETVPADLAERLRQVRSNGFSDMSLQTAGVRNISAPILTLDGTALAVI
TCCPYITLGGTAPTRETCEQLIRDAAKRISEVVTGNV"

/complement(972..1937)

/gene="AGR_L_812"

/complement(972..1937)

/gene="AGR_L_812"

/note="transaldolase PA2796 {imported} - Pseudomonas

aeruginosa (strain PA01)"

/codon_start=1

/transl_table=11

/product="AGR_L_812p"

/protein_id="AAK88978.1"

/db_xref="GI:15158762"

/translation="MTSKLEQLRAITVVADTGDIENAVAREKPVDCNTNPTVLKALG
TPAFADAVKEAVWVGKGGQSDAVVAADVADRLAISVGAALAGLVGGRVSTVEDADLS
FDTQASITKARHIIASIKYKGERIERILIKLASTWEGIKAAEVLOSIEDCNLTLLFS
QAQATACAEAKAFILSPFVRILDMYKKSTGETYTAETDPGVSVRSIYKANGIG
TVMGASFRNVGEIEALACDRLTISPALLELDKDKGLERKLSQDNVYKAEALQSLD
EKAFRWANNEDAMATEKLSERLFAKDLVTLRENVKELTAAAA"

2133..3164

/gene="AGR_L_814"

2133..3164

/gene="AGR_L_814"

/note="(AP001509) transcriptional regulator"

/codon_start=1

/transl_table=11

/product="AGR_L_814p"

/protein_id="AAK88979.1"

/db_xref="GI:15158763"

/translation="MAKTVAKLRRETHHTYSEASLRLRAAMLYYNOGMTQKDVAREL
GISRSTVIRLDEAMKRSVQIWIWEGIEDFVSLAGOLEAAVGLDEAVTIIPSPGPSA
GPSTASPTASKTSAEGTAKAVGLAQGLFSEVVPNGATIGVWGRWTATSLSSRPP
RRENCVVSLGGIVAVQPLDPTWRLASALGAECYMFAPLPLVDSVETKRALIEK
CGLATYDLAETDLAIYSCGDIPGHSTLSGEGTSRETLRELVDAGVCVDFMFMFID
AEGRSVDHPINORAMAIDTLRKAKHIVLASGGAAHRAIATIRATIKRIGCNTLTIDEA
AARALMELV"

/complement(3161..3448)

/gene="AGR_L_815"

/complement(3161..3448)

/gene="AGR_L_815"

/note="HYPOTHETICAL 11.0 K0 PROTEIN Y4Kp"

/codon_start=1

/transl_table=11

/product="AGR_L_815p"

/protein_id="AAK88980.1"


```
/db_xref="GI:15158764"
/translation="MRLIWRVRLRELDALGVYTAERNPRAAKVVRVRAIHQTTARLLS
DNPLGRSGIEGTRRLVPLVPLVIVAYRTEQVDIVFVQHAAREWPDPA"
complement(3454..3690)
/gene="AGR_L_816"
complement(3454..3690)
/gene="AGR_L_816"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_816p"
/protein_id="AAK88981.1"
/db_xref="GI:15158765"
/translation="MKSTIELSDDKRRLDLAERSNSTPSRIEDALSHGRSLAWOE
KWTSGVAGLAERADAGEFVEEINDLVNKVAKA"
complement(3743..4615)
/gene="AGR_L_818"
complement(3743..4615)
/gene="AGR_L_818"
/feature="O6-methylguanine-DNA methyltransferase PA2118
(imported) - pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_818p"
/protein_id="AAK88982.1"
/db_xref="GI:15158766"
/translation="MNANIMLNEDITPIGSDYDVRGVIELTLDYREOPSLEAIAAR
LGOSPTOLQFTFRWAGLSKPAFLQAVTLQHAKEELLREEDLPLETSIEVGMSPGRL
HLDFVTEAMSPGEWAKAGGLITRYGFHSPAGLAVNITDRGLACGADGERA
CFDEAMGNADYVEDREATAFYAARIFEPAMWTADKPLRVLLGDFQVRWKSLL
KIPMGRAVYTSNACDIOPTASRAVGAAGVANSFVVPCHRAVKGSGALTGYHWWGL
TRKRAMLGWETGKA"
complement(4882..5364)
/gene="AGR_L_819"
complement(4882..5364)
/gene="AGR_L_819"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_819p"
/protein_id="AAK88983.1"
/db_xref="GI:15158767"
/translation="MDTINDOPIFAEELVPHRSIGRGRERLLLLTGFACLIYGGFEL
ATGAMPILGFLFGDLFLLYIAFRANYRAAKAEVSVSRNLSIRKFSAPGRMVVEYRF
NPFAWRVRVRHDEIGIVSMHVTGEGRATIDIGSLPNLDDRESFAKAFGALATVRRRM
"
5402..6184
/gene="AGR_L_820"
5402..6184
/gene="AGR_L_820"
/feature="endonuclease III PA3495 (imported) - pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_820p"
/protein_id="AAK88984.1"
/db_xref="GI:15158768"
/translation="MTVAKKRSSTQTLKKSIPAORRKPVRVKTAYSKDELTEIFRRS
IQREPGELEHTNPTLLVVALSAQATDVGNRATRALFKVADTPEKMLALGEEQL
IGHKLTIGLVNRKANVIALSQMLIDNFGVEVPKTRBELVTLPGVGRKRTANVMSMAF
GVPTLADTHVFRANRLCLAPKGTITDEVEDLVRIIPEOYLPHAHHLILHGRYCK
ARKPECRCVIADICKSPEKTFDIPAPLVELPAQLFGAAGE"
complement(6146..6883)
/gene="AGR_L_822"
complement(6146..6883)
/gene="AGR_L_822"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_822p"
/protein_id="AAK88985.1"
/db_xref="GI:15158769"
/translation="MIFDAARLAFGNLFAAETRSVFWKVLGLTLLVLAALWFAIRSF
```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```
IYFALPWIDTLIPGVPDWAGWLSFVFAIFAGIGIALALALLISPTVAVIAGLFLDDVA
DYVEKONYPDPGKAMPIGEAVLSSIKFFGVVIAGNLVALLLLVPGVNLIAFELVN
GYLIGREFFEFAPARRFSPADARLFRVYKHRTVTVMFAGLVIAAFLAVFLNLITPLEAA
SIWVHLHKAYSRDPSPFAAGTEQLRG"
7179..8012
/gene="AGR_L_824"
7179..8012
/gene="AGR_L_824"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_824p"
/protein_id="AAK88986.1"
/db_xref="GI:15158770"
/translation="MVEKLLNLNTRAMKAGTPSPERKIATYLIHLEDELPPETAQTLAA
KLSLSPMVTGREFLSLGYROFSDIRADLRHAETVGDOPVTAEGGERSHSPFSQILLS
QQIQAIQTADMTTQTWRLAMNEITSADDOVCLATSPGEGLDAGRHFGYRLLCEGRGHIH
YLGSDNAYVALWDYDIPARTLLVIMDCGGNLLPQRLSTARKGGYRLLITTRFYEW
GPESVDLCAMPQSQNGGQLQLVLSLEFTLCALSAGADGAKRVKNLAALKRSLR
S"
complement(8117..9694)
/gene="AGR_L_826"
complement(8117..9694)
/gene="AGR_L_826"
/feature="D-ribulokinase (EC 2.7.1.47) - Klebsiella
pneumoniae"
/codon_start=1
/transl_table=11
/product="AGR_L_826p"
/protein_id="AAK88987.1"
/db_xref="GI:15158771"
/translation="MMRQNLVAVDVGTASARAGIFDPAGRLIARSIHPLMQRPRENH
AREHSDTIWNAVCIAKAAALADAGVLPQSIAGIGDATCSLVIRDERGEPVSVSTDD
DREDTIVLQHRAIGEADRLTASGRHVLDFAGNSVSPQMOKMLWKTMPQSKSRM
SFAEDLADELTKATGSAQRNSCTQTAKWNLAEHFGWQADYLAFLGDLDLKERAGL
PETTVMEFGAIGPLUSPAEELGDTGCQVAAQMDIDATGALGALGGCLAEDEGKHA
LIAGTSSCLVAMSTQPMPSLWGPQWAVLPQHWLVVEGGSATGALLDHIVRMHAAG
GEPDTALHARIVARVTELRELEGFADRLHLPDFHGNRSPADPHAVGVVSGILTD
TSFDSLCLRLYRTAIALGARHVLDMAREFGYAVETLHVTGGHVKNPLLMELYADV
GKRVIVPATADAVLGLTAMTAAATAGGVHASLAAAGAAAMPYPCNAETISGNPALAAHYERD
```

Query Match 6.1%; Score 112; DB 1; Length 14945;

Best Local Similarity 48.1%; Pred. No. 4.5e-23;

Matches 415; Conservative 0; Mismatches 433; Indels 12; Gaps 3;

QY 742 GATTTCACAAACGCTCATGTCTGAAGGGCAGATCCGAAGACAAAAGTAGTTCCTTCCACC 801

Db 11629 GATACCCAGGCCATCATCGATAGCGTTGCCGATCCGAAAATCCGATCGTTTCGATGACG 11570

QY 802 GTACCCGAAGGGGGTACTACTGTAATCCAGCTACAACTCGAAATTAACAATCCTGAT 861

Db 11569 ATTACCGAGGGGGTTATTTCATCGATGCTCCGCTCCTTCAATCCGAGCATCGGGCC 11510

QY 862 TTAGCGGCAGATCTTAAAGGGGGATGC-----AAACAAATTTACGGTGTATTACCCGT 915

Db 11509 ATTTCGGAAGACGGCAAGAACCCGCTCCGATCCGAAAACCGTGTTCGGTTTCATGTTGC 11450

QY 916 ATCCTCGAAGCGCGTATGCAATAAACCGGACCACTAACCCCTGCTGAATTCGCGATAAC 975

Db 11449 GGGCTGAAGGGCGGGCGGATGAAGACCTTCAGCCCTTCACCGTATGTCCTCGCGACAAC 11390

QY 976 GTGCCCAATAATGTGAACGTTTCCATGATGCGCTTGGTGTTCCTCCAGCATTAACGTGC 1035

Db 11389 ATTCCCAACATGGCAAGGTGACGAAACGGGTGGTGGCGCTGGCAGCGCTCTC--C 11333

QY 1036 AAACAGGATGTCATCGACTGGCTGAGTACAAATACACACTTGCCTCCGCAATACCATGGTTCAC 1095

Db 11332 GACCCGGGCTTTGCCAACTGGATCGCGGAAAATATGCGCTTTCCGAAATTCATGTCGAC 11273

QY 1096 CGCATTCACGCTCGTCCGGCAGCAGAACTTCGGGACCGGATCAAGCTCAACGGGTATT 1155

Db 11272 CGCATCACCCCGCCACCGGGCGGCGTGAACGCAATATCGCCCGTGACGATTCGCGCATC 11213


```

LGDILHLIRACALNFGAGPRGGAGDEEDRSITNEEPIIPSVDEHGLKVKCLRSNPNT
RLRLKTLDAVKALLVSCCARTARDLIDFDNNGVAMMKWIKILYHEVAQETTLKDSYR
ITLVPSSDGISVCGKLGWELNCFHMPASFVIEKFCSENSELTSLWTKILFFK"
2393..3004
/gene="YEL076W-C"
CDS
2393..3004
/gene="YEL076W-C"
/codon_start=1
/evidence=not_experimental
/product="Yel076w-cp"
/protein_id="AAB65035.1"
/db_xref="GI:1857035"
/translation="MGSSFVIDRSSSPAPRGPAPKLSAHARKIICKISPNRSLFLFI
ISLHCIKYFESLGRHSRFSVSSTFFSPGLACIAHLRVGQIVPIFYGAFLK
TPYRNCAGNKVSESTRHRAVVRPSTRVFTVQDETOLIIIVSSVEVKKRGIVILSI
EFQSMHLKORVDHVVCLGLVRPGNLVRLQQTSGTFLHESRP"
complement(2868..3236)
/gene="YEL075C"
CDS
complement(2868..3236)
/gene="YEL075C"
/annotation="similar to hypothetical 65.4 kD protein in
subtelomeric V' repeat from S. cerevisiae, Swissprot
Accession Number P24088"
/codon_start=1
/evidence=not_experimental
/product="Yel075cp"
/protein_id="AAB65012.1"
/db_xref="GI:603243"
/translation="MKVSDRKEFEKANDEFESALNNKNDLVHCPISITLFSIPEYVR
SFYEDKSLGLKVKYFRGMDRKRSEKVVISVWGNVKKFLTFVEDEPDFGGPI
PSNRPDGLHVSSAYFEIQ"
3649..3987
/gene="YEL074W"
CDS
3649..3987
/gene="YEL074W"
/codon_start=1
/evidence=not_experimental
/product="Yel074wp"
/protein_id="AAB65013.1"
/db_xref="GI:603244"
/translation="MDLRSALKETNIKSKLHLFIRVTTSRFEVKEPFLAYLHMARQ
KDKLLPKCITPPNKKADPFTNDVSSYFTHTHPCSPSLVYTYVNTTEKSPSKPKH
KNILFPNFK"
complement(4753..5076)
/gene="YEL073C"
CDS
complement(4753..5076)
/gene="YEL073C"
/codon_start=1
/evidence=not_experimental
/product="Yel073cp"
/protein_id="AAB65014.1"
/db_xref="GI:603245"
/translation="MVNLANVLTNATATLSAWSNTVPLETYFHFDEASGFDYVLNV
SVIWMNETLYETRIVPALINVRWLDHMEANDPSPSVTNPYETSGYAFSVVPLMG
NMKVA"
11243..11938
/gene="YEL072W"
CDS
11243..11938
/gene="YEL072W"
/codon_start=1
/evidence=not_experimental
/product="Yel072wp"
/protein_id="AAB65015.1"
/db_xref="GI:603246"
/translation="MSACPNIIVILPVEILNKKSKDTKYSLYTTINRGYDVPRLKYGI
IVSPRVHSLFTLSDLGFDKNIEKSSLYLLNDPLAYPNFHEHFEQURKGTNKLDSL
PTYIYKQVTEAPDESHLTATIGYRPNKSEYEITGTSMNGYKGLKFNYSVIHM
MRSHKCRKRVADIIMEHDLGLGYEKKLGFVEVQRFKLKEHQVQKVFDDKDFH
VIMIKELGNHRL"
13878..15368
/gene="YEL071W"
CDS
13878..15368
/gene="YEL071W"

```

```

/note="similar to D-lactate dehydrogenase from S.
cerevisiae, Swissprot Accession Number P32891"
/codon_start=1
/evidence=not_experimental
/product="Yel071wp"
/protein_id="AAB65016.1"
/db_xref="GI:603247"
/translation="MTAAHPVAQLTAETAYPKVRNPFKVLDSLAYFRSILSNDIEI
LNSQPEELASFNDQNMKKYRGQSNLILPNSTDKVSKIMKYCNCKKLAVVPOGNTD
LVGASVPVFEIVLSLRNMKNRQDFDPSVGTFCDAGVVMRDAHGFLLDHDHIFPLDL
PSRNNQVGVVSTNAGLNFELRYGLSHGNVLGLEVLNPGELIISINAKLRDNTGYD
LKQLFGAEGTIGVVTGIVSIVAARPKALNAVFFGIENFDVQKLFVAKKSELSEILS
AFEMDRGSIETIEYLDKLPFPLENQHNHYVLIETSGSNKRHDDDEKLTAFKLTDTDS
KLISEGMAKADKADRLWTKRVSPTACNSYGYMKYDMSLQDLKDLVSVAATVERL
NAAGLIGDAPVVKSCYGHVGDGNIHLNAVREFTKQIEDLLLEPFPVVEYIASKKGS
ISAEHIGIFHKKGLHYTRSDIEIRFMKDINKHDPNGLNLPYKI"
Query Match 6.0%; Score 111.4; DB 8; Length 50849;
Best Local Similarity 49.6%; Pred. No. 8.8e-23;
Matches 314; Conservative 0; Mismatches 316; Indels 3; Gaps 1;
QY 949 CCACCTAACCCCTGCTGAATTGCGATAACGTCGCGCCATAATGGTGAACGTTTCCATGATGCG 1008
DB 17682 CCATTACATATTATGTCATGTGACAAACATGCCCCAAAATGGTGTACAGTAAGACCATG 17741
QY 1009 CTGGTTGAGTTTCTCCAGCTAACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAAT 1068
DB 17742 CTTGTTGCAATTTGCCAAGTTAAAGAGGATGAGAATTCGCCGCTGGATTGAAGACAAG 17801
QY 1069 ACCACTTGGCCGAATACCATGTTGACCGCATTCACGCTCGTCCGCGCAGCAGAACTCCG 1128
DB 17802 GTTACTTCTCTAACAGCATGTTGGACCGTGTGACCCACGTTGTACCGATAAAGACGCT 17861
QY 1129 GCACGGATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGGAAACCTTT 1188
DB 17862 AAATACGTTGCTGACACCTGGGGAATCAAGATCAATGTCCCGTTGTCGACGAACTTTC 17921
QY 1189 ATCCAGTGGTCTGGAAGATAAATTCGCTGATGTCCTCGCGCACTGGAGAGGTCGCT 1248
DB 17922 ATCCAATGGTCTCTTGAAGACAACTTCTCCGATGGCGCTCCCTCCACGAACTTGTGCT 17981
QY 1249 GTCGAACCTGGTGGCGTCGTAATCCCTATAGAGAGCGCAAGATTGCGATTCTTAATCT 1308
DB 17982 GTTCAGGTCGTCAGGATGTCGATTCCTACGAATTTGAAGTTGCGTCTACTTAACGCT 18041
QY 1309 TCACACAGTTGCATCGCTGGCGAGGTACGTTAATCGGTCAAAAATATATATCCACGAAAGC 1368
DB 18042 GGACATTCGCTATGGGATATTGGGATATTGGGACGCTACACCTATATACATGAGTT 18101
QY 1369 ACAATGACCGATTTTATATCATGATTGCCGACCGCTAGCTGACGGAAGATGTCATTCT 1428
DB 18102 GTCACGACGCCAACTATCAACAAGTATATCCGTTGTTTGTATGCGTGAGGAAGTTATCCCA 18161
QY 1429 TCGTTG---GGCGTAACCGTATCGATTGGCCAACTACCGTGATGTTGTACTCAACGCT 1485
DB 18162 TTATTCCTTAAAGTGGCAGGTGTTGATTTTCGAAGAGTACACTGCATCATGTTGGGAAGA 18221
QY 1486 TTATCAATCCACATATTCAGGACACCAACCAACGCGTCGCTCGCATGTTCTCTCGAA 1545
DB 18222 TTCTCCAATCCAGCAATTCAGGACACTGTGCGACATGTTGTTGTTGATGGCTCTGTTAAG 18281
QY 1546 ATTCGCGGCGATGATTGCCGCCCACTGCAGAG 1578
DB 18282 ATGCCTAAGTATGTTTGGCCATCAATTTACGAG 18314
LOCUS AE005449 10449 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 68
of 290.
ACCESSION AE005449 AE005174
VERSION AE005449.1 GI:12516490

```

KEYWORDS	Escherichia coli O157:H7 EDL933.	gene	complement(2369..4060)
SOURCE	Escherichia coli O157:H7 EDL933		/gene="fruA"
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia	CDS	/note="23425"
REFERENCE	1 (bases 1 to 10449)		complement(2369..4060)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		/gene="fruA"
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7		/function="regulator; Degradation of small molecules:
JOURNAL	Nature 409 (6819), 529-533 (2001)		Carbon compounds"
MEDLINE	21074935		/note="Residues 1 to 563 of 563 are 99.46 pct identical to MG1655: B2167"
PUBMED	11206551		/codon_start=1
REFERENCE	2 (bases 1 to 10449)		/transl_table=11
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		/product="PTS system, fructose-specific transport protein"
TITLE	Direct Submission		/protein_id="AAG57305.1"
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		/db_xref="GI:12516493"
FEATURES	Location/Qualifiers		/translation="MKTLLIIDANLGOARAYNAKTLGAAARKAKLEIIDNPDAEMA
Source	1..10449		IVLGDSPNDSALNGKNWNLGDISRAVAHPLELSEAKGHAKPYTPAATAPVAASG
	/organism="Escherichia coli O157:H7 EDL933"		PRKVAATACPTGVATFMAAEIATEAKRGWVKVTRGSGAGNATLPEVAARD
	/strain="EDL933"		LIVIAADIEVDLAKFAGRPWYRTSLGALKTKAQELDKAVAEATPEPAKAQTATSE
	/serotype="O157:H7"		GNESAGAYRHLLTGVSYMLPMVAVAGGLICIALSFAGFEAFKEPTGLAAQIQGGS
	/db_xref="taxon:155864"		AFALMPVVLAGYIAFSIADRPGLTGLIGMLAVSTGSGFIGGIIAGFLAGYIAKLIS
gene	/note="enterohaemorrhagic"		TOLKLPQSMALKPILIIPLISSVLVGLAMIVLIGKVPVAAIIEGLTHWQWTANAV
	complement(78..1016)		LGATILGMMCTDMGSPVNAKAAVFGVGLLSTQTYGPMAAIMAAVGPLANGATMV
	/gene="yeiN"		ARRKEDKAOEGGKAALVLCFISEGAIFFAARDPMRVLPCCIVGGALTGAISMAIG
	/note="23422"		AKLMAPHGLFLVLLIPGATIPVLGYLVAIIAGTLVAGLAYFLKRPEDVAKAA"
	complement(78..1016)		complement(4077..5015)
CDS	/gene="yeiN"		/gene="fruk"
	/function="orf; Unknown function"		/note="23426"
	/note="Residues 1 to 312 of 312 are 99.35 pct identical to MG1655: B2165"		complement(4077..5015)
	/codon_start=1		/gene="fruk"
	/transl_table=11		/function="enzyme; Energy metabolism, carbon: Glycolysis"
	/product="orf, hypothetical protein"		/note="Residues 1 to 312 of 312 are 100.00 pct identical to Strain MG1655: B2168"
	/protein_id="AAG57303.1"		/codon_start=1
	/db_xref="GI:12516491"		/transl_table=11
	/translation="MSELKISPPELLQISPEVODALKNPVVALESTIISHGMPFPQN		/product="fructose-1-phosphate kinase"
	ACTATEVEETIRKQAVPATIAIIGVMKVGLSKEIELELGRGHNVTKVSRDLPEV		/protein_id="AAG57306.1"
	VAAKNGATTVASTMIIAALACIKVPATGGIGGVIRGAHEFTDISADQLANTVTV		/db_xref="GI:12516494"
	VCAGAKSLDLGLTTEYLETGPVLIGTQKALPAFFCFTSPFDVSIRLDSASEIARA		/translation="MSRRVATITLNPAYDLVGFCEPGEVNLVKTTLGLHAAGKGIN
	MAVKWQSGLGLVAVNPPEGFAMPEHTINAVIDQVAEAEAQGVIGKESTPFLAR		VAKVLKDLGIDVTVGGLKQNDGFQOLFSGLIANRFOVVQGRTRINVKLTEDGE
	VAELTGGDSLSKNIQLVFNAILASEIAKEYQRLAG"		VTDNFPSGFETPADWERVTDLSLWGFQDMVCSVSGSLSPVSPFAFTDMWTRLRSQ
gene	complement(1004..1945)		CPCIIFDSSREALVAGLKAAPWLVPNRRELEIMWAGRKLPENKMDVTEAAHAREQGIA
	/gene="yeiC"		HVVISLGEALWVNASGEWIAKPPSDVSTVGAGDKSLVGLLIYGLLMRESSEHLR
	/note="23423"		LATAVAALAVSQSNVGTIDRPQLAAMMARVDLPFN"
	complement(1004..1945)		complement(5015..6145)
CDS	/gene="yeiC"		/gene="frub"
	/function="putative enzyme; Not classified"		complement(5015..6145)
	/note="Residues 1 to 313 of 313 are 99.04 pct identical to MG1655: B2166"		/gene="frub"
	/codon_start=1		/function="enzyme; Transport of small molecules:
	/transl_table=11		carbohydrates, organic acids, alcohols"
	/product="putative kinase"		/note="Residues 1 to 376 of 376 are 99.73 pct identical to MG1655: B2169"
	/protein_id="AAG57304.1"		/codon_start=1
	/db_xref="GI:12516492"		/transl_table=11
	/translation="WREKDYVVIIGSANIDVAGYSHESLYNADSNPGKIKFTPGVGR		/product="PTS system, fructose-specific IIA/fpr component"
	NTQAQLNLGNKWLLSAVGSDYFQSLTLQNSGVVYDKCLIVPGENTSYLSLLD		/protein_id="AAG57307.1"
	NTGEMLVANDNINAITAEYLAQHREFIQRAKVIADCNISEEALWLDNAANVP		/db_xref="GI:12516495"
	VFVDPVSAWKVKVRLRSQIHTLKNRLEAETLSGLASGREDYAKVAWPHQHLGN		/translation="WFQLSVODIHPEKAGDKKEAIRVOVAAALVOAGNVAEGVNGL
	RLVLSMGDGVYSDISGSGSWAPIKTNVINVTGAGDAMMAGLASCWVDGMPFAESV		AREQOTSFLGNGIALPHGTTTDRQVLTKTGVOVFQFPEGVTWGCQVAYVAINAAS
	RFAGQCSSMALSCEVTNNPDLSIANVISLVENAECLN"		ASDLTLQALNAARLKEAGVADATVTKAINEQPLNQGQIWLSDSAEGNLRSATVX
			RAANAFDVGDETAAMLVSVAMNDQPIAVLKRLDLLLDNKRLLIKADAATLLALT
			SDDAPTDDVLSAEFVVRNEHGLHARPGTMLVNTIKQFNSDITVTNLDGTCGKPANGRSL
			MKVVALGVKKHGRHLRFTAQGADAEQALKAIGDAIAAGLGEGA"
			6513..7694
			/gene="yeiO"
			/note="23428"
			6513..7694
			/gene="yeiO"
			/function="orf; Not classified"

JOURNAL

Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kengen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)

COMMENT

genome project.

FEATURES

Location/Qualifiers

1. .268857

/organism="Escherichia coli" O157:H7"

/strain="O157:H7"

/sub_strain="RIMD 050952"

/db_xref="taxon:83334"

complement(73. .2271)

/gene="ECs2865"

complement(73. .2271)

/gene="ECs2865"

/note="similar to B2060_ECOLI gii1788374 percent identity

99 in 732 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="hypothetical protein"

/protein_id="BAB36288.1"

/db_xref="GI:13362334"

/translation="MGAGIERRAGIRMTKVKQHAAPVTGSEIDIGRLVGTVEARW
WIGITAVFALCAVYTFATPIYSADALVOIEQSGNSLVODIGSALANKPPASDAE
IQILRSVLVTDDLDIDIAVSNTFPIFGAGWDRMQNQNTKVTNRPKEME
DOQFTLVNDKNYTLSSDGGFSARGAQOILKEGVTLMVEAIHARPGSEFTVKYS
TLGMINQLQNTLVTEGDKAGVLSITYTGEDREOIRDLNSIARYOPONTERKSAE
ASKSLAFLAQLPEVRSRLDVAENKLNAPRQDKSDVLPLEAKAVLDSMNVNDAQNE
LTFKAEISKLTVPAYPTLLEKRAQLEDEKALNGRVAMPKQIEVRLTRDVE
SQQYVMOLLNKEQKLTAEATSTGVDRIVDPAITOPGVLPKPKLIILGAILGLML
STVGLVLSFRGSESPVLEEHSYASIPLSSEWQKARDSVKTIKIKRYKQSOL
LAVGNPTDLAIEAIRSLRSLSHFAMQAOANNVMTGYSPSTGTFVCANLAAVISOT
NKRVLIDCMRKGTHYELLLGNTNNVGLSEIILGODITTAAKPTSIKAFDILPRGOV
PNPSELLMSERAEFLVNWASKNYDLVILDTPEILAVTDAALVGRHVGCTLMVARYAV
NTLKEVETSLSRFEONGIPVKGVILNSIFRRASATQDGYEIEYKDSNAK"

complement(2238. .2681)

/gene="ECs2866"

complement(2238. .2681)

/gene="ECs2866"

/note="similar to WZB_ECOLI gii1788375 percent identity

100 in 147 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="probable protein-tyrosine-phosphatase"

/protein_id="BAB36289.1"

/db_xref="GI:13362335"

/translation="WFNNILVVCVGNICRSPTEARLLQRYHPHLPKVESAGLALVKKG
ADPTAISVAAEHQSLGHCARQISRLCRNYDLILTWEKRHIERLCENAPEMRGKVM
LFGHWDNECEIPDPYRKSRETPAVYTLLEARSARQWAQALNAEQV"

complement(2687. .3826)

/gene="ECs2867"

complement(2687. .3826)

/gene="ECs2867"

/note="similar to WZA_ECOLI gii1788376 percent identity

100 in 379 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="putative polysaccharide export protein"

/protein_id="BAB36290.1"

/db_xref="GI:13362336"

/translation="WMKSKMLPLLVSTLSGCTVLPGSNMSTMKGDKVIKQDADF
DLDKMWNYPILTPRLIDQLRPNRVARNMTLSEIANQYRVGPDVNLVNTVWHPH
LTPTAGVRSRSDKYNWQVPGDTGTFPYIGKVHVVGKTLAEIRSDTIGRLATYIADPQ
VDVNTAIFRSKDAYTSQVNGSQQAITNPLTILDAINAGGLDITADWRNVLTIN
GREERISQAELQMDTDLNCRLLYPGDILYVPRNDLKVFMVGEVKKQSTLKMDFSGM
TLTEALGNAEGIDMTNSNAGSIFVIRPLKGEGRNGKIANIYQDLMSDATSLVMATEF
RLQPDVVYVVTAPYSRWNRLLNQLLPTISGVRYMTDASDIHNH"

4419. .6068

/gene="ECs2868"

CDS

4419. .6068
/gene="ECs2868"
/note="similar to YEGH_ECOLI gii1788377 percent identity
99 in 549 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein"
/protein_id="BAB36291.1"
/db_xref="GI:13362337"
/translation="MPRFIDRIVTFVIOIGFVLTWMIADPSIWAAGLITLIVIELV
LGIDNLVFIATLAEKFLPKQDRARVTGLLAKRLLLSASISWLVLTLPQLFSFRS
FTSARDLIMLFGFFFLFKATMELNERLEGKDSNNPTQRKGAFFGVVTVQIVLDAI
FSLDSVITAVGMVDHLLVMAAVVIAISLMASKPLQTFVNSHPTIVILCSFLMLI
GFSLVAEGFVPGYLYAAGFCFVMIQALQALFNRRFLSANQTLRQRTTEAVM
RLLSQKDAEALDASTMLVDHNOQIENPOERRMIERVNLNORTVSSIMTSRHDI
EHIIDNAPDEIRQLLERNQHTRLVVTGDDAEDLLGVHVVDLLOQSILRGEPLNLRV
LIRQLVPFETPLLPALPQFRNARHFAFVVDGFGSEGVITLSVTETIAGNLNE
VEEIDARHDIQKADGSWTANGHMPLDQVYVPLPDKREYHTTAGLMLLEYLQRP
KPEEYQVGDYLLKTLQVESHRVQKVIILPKRKDGEMEYEV"
6143. .6481
/gene="ECs2869"
6143. .6481
/gene="ECs2869"
/note="unknown, similar to hypothetical proteins e.g.
gii19106618|gb|AAF84382.1|AE003986_12 [Xylella fastidiosae]
percent identity 60 in 105 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB36292.1"
/db_xref="GI:13362338"
/translation="MRKLAFLDTSDDLARPESSROEIGYQLDRIQOGLNPDWKP
FSTIGPGVREIIRTDADGIYRVYAKFEAYVVLHCFQKQTTSQSDIDLAKRYK
ELVQERKNEN"
6471. .6761
/gene="ECs2870"
6471. .6761
/gene="ECs2870"
/note="unknown, similar to hypothetical protein {Xylella
fastidiosae} gb|AAF8486.1|AE003993_5 percent identity 52
in 86 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB36293.1"
/db_xref="GI:13362339"
/translation="MKIETFDVWDVAVSDTPQOAEHNRIRAEVLVTIINWIEQGFSG
AQASALGVTPQRISELARGKIQFISDKLITMMAHAGLHIQRIEYQYPHAA"
complement(6814. .8667)
/gene="ECs2871"
complement(6814. .8667)
/gene="ECs2871"
complement(6814. .8667)
/note="similar to ASMA_ECOLI gii1788378 percent identity
99 in 617 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="suppressor of ompF assembly mutants"
/protein_id="BAB36294.1"
/db_xref="GI:13362340"
/translation="MRRFLTMLLVVLVAGLSALVLLVNPNDRDYMWKQVAAARSG
YQLDGLPLRWHPQLSILSGRMSLTAGASQPLVRADNRMLDVALPLSHQLSVK
QVMLKAVIQLTPQAEVSDAPVAPRNTLPDSDDRGWSFDSLSKLVADSVLVFQ
HDDQCVTLIRNLQMEQDQHRGSEFSGRVNRQDRDLTISLNGTVDSDYPHDLTA
ATEQNLWQGLADLPKQIQGQSFQAQSGKRKLSFNQISLNTANDSTLSGQAQVTL
TEKBPWLQRLQPOLNLDNLIPNETANGCAAGQOQSOTLPRVIVSRISDEPAYO
GLOQTADILLQASNRVRGMNFTVATQNTKSGILLEITQGLKNGQVSLPGTLD
ATSNIPRNFQRLNVEIGITLKAENYFISLTGKMSLAGDSGADIDADARHNQOG
QAHVEADTRMEGMNFQMIQQAQVDRKAAENFNDVRLDRFTTDLTKDGVVT
LNDMQGQSPVLTALTEGGMNLNADQCTQDFDIRVVGWNGESKGLDFLTKPVLPRVY

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:12:47 ; Search time 409.996 Seconds
(without alignments)
10150.564 Million cell updates/sec

Title: US-09-802-208B-1

Perfect score: 1848

Sequence: 1 atgaacgaacattacatg.....ttacgcgttaactaa 1848

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1848	100.0	1848	22	AAD16810 Escherichia coli s
2	105	5.7	4115	21	AAA97430 Gluconobacter oxyd
3	101.8	5.5	1458	20	AAZ35672 Gluconobacter subo
4	87.6	4.7	9810	20	AAZ32025 Human METH1 relate
5	87.6	4.7	9810	22	AAC90082 AF018073 cDNA clon
6	80.6	4.4	921	23	AAZ92759 DNA encoding novel
7	73.4	4.0	2810	23	AAZ78056 DNA encoding novel
8	68.8	3.7	1509	22	AAH65091 C glutamicum codin
9	68.8	3.7	1632	22	AAF71527 Corynebacterium gl

c	10	68.8	3.7	349980	22	AAH64966 C glutamicum codin
c	11	66.8	3.6	1182	23	AAZ93068 DNA encoding novel
c	12	61.2	3.3	684	21	AAF12753 Aspergillus oryzae
c	13	43.4	2.3	2193	23	AAZ82343 DNA encoding novel
c	14	42.4	2.3	14041	22	AAH48024 Internal control B
c	15	38.4	2.1	2238	23	AAZ92760 DNA encoding novel
c	16	33.2	1.8	278	22	AAO10154 Human reproductive
c	17	33.2	1.8	278	23	ABL96521 Human testicular a
c	18	33.2	1.8	4787	20	AAZ24631 Human lung tumor a
c	19	33.2	1.8	4797	21	AAZ65870 Human lung cancer-
c	20	33.2	1.8	4797	24	ABN97444 Gene #3942 used to
c	21	33.2	1.8	4797	24	ABL68585 Kidney cancer rela
c	22	33.2	1.8	4797	24	ABL49089 Human lung tumor
c	23	32.8	1.8	615	23	AAZ51468 Pseudomonas aerugi
c	24	32.4	1.8	2230	24	ABL56160 Maize acyl CoA oxi
c	25	32.4	1.8	2436	23	AAZ54141 Pseudomonas aerugi
c	26	32.4	1.8	3098	23	ABL05639 Drosophila melanog
c	27	32.4	1.8	3271	24	AAI67053 D. melanogaster PA
c	28	32.4	1.8	5166	23	ABL05638 Drosophila melanog
c	29	32.4	1.8	134141	24	ABN83487 Escherichia coli O
c	30	32	1.7	512	24	ABN16299 Salmonella typhi D
c	31	32	1.7	2139	23	AAZ56324 Human ORFX polynuc
c	32	32	1.7	3725	21	AAZ99500 DNA encoding a mai
c	33	32	1.7	3725	21	AAZ99506 DNA encoding a mai
c	34	32	1.7	3725	21	AAZ99521 DNA encoding a mai
c	35	31.8	1.7	2022	23	ABL09849 Drosophila melanog
c	36	31.8	1.7	2061	21	AAZ76801 Human ORFX ORF2356
c	37	31.8	1.7	2083	22	ABA09036 Human PTD014 homol
c	38	31.8	1.7	2084	22	AAO6020 Human reproductive
c	39	31.8	1.7	2084	22	AAO6021 Human reproductive
c	40	31.8	1.7	2084	23	ABL98585 Human testicular a
c	41	31.8	1.7	2084	23	ABL98586 Human testicular a
c	42	31.8	1.7	2977	23	ABL17060 Drosophila melanog
c	43	31.8	1.7	4022	23	ABL09848 Drosophila melanog
c	44	31.8	1.7	23620	23	ABL29054 Drosophila melanog
c	45	31.8	1.7	23620	23	ABL19102 Drosophila melanog

ALIGNMENTS

RESULT 1

AAD16810
ID AAD16810 standard; DNA: 1848 BP.

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

XX AC AAD16810;

e.g., arabinol, ribitol which confer selective advantage on transformed cells
 Claim 1; Page 34-35; 37pp; English.
 The present invention relates to a positive selection system that involves conferring to transferred cells the ability to metabolise arabinol, ribitol and/or mannitol. The positive selection method is used in positively selecting transgenic cells from a population of cells using the positive selection method, the presence of the gene of interest in the genetically transformed cells may be determined without the disadvantages associated with traditional negative selection systems. Positive selection of the transformed cells is achieved without directly damaging the neighbouring non-transformed cells. The transformed cells may be identified by simple visual means without the use of a separate assay to determine the presence of a marker gene. This technique also avoids the release of antibiotics or other dangerous genes into the environment. The present selection is *Escherichia coli* strain C arabinol dehydrogenase gene.
 Sequence 1848 BP; 494 A; 443 C; 473 G; 438 T; 0 other;
 Query Match 100.0%; Score 1848; DB 22; Length 1848;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1848; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 QY 1 ATGAACGAACAATTTACATCGCTGCACATCGGTTAGGTTCTTTTCATCGCGCACATCAG 60
 DB 1 ATGAACGAACAATTTACATCGCTGCACATCGGTTAGGTTCTTTTCATCGCGCACATCAG 60
 QY 61 GCGTGGTATCTACACCGTTTTCAGGTGATGGGGGATAAAGCGCTGGAGCATTTGCTGCGGGC 120
 DB 61 GCGTGGTATCTACACCGTTTTCAGGTGATGGGGGATAAAGCGCTGGAGCATTTGCTGCGGGC 120
 QY 121 AATATTGTAATGATGCTGAACATGCTGACAGGCACTAGTCCACAGAAAGGTCGCTAT 180
 DB 121 AATATTGTAATGATGCTGAACATGCTGACAGGCACTAGTCCACAGAAAGGTCGCTAT 180
 QY 181 GTGCTGGAACCGTACGCCGGAAGGGTAAGCGAATATGAACAGATCACCTCAATTCAG 240
 DB 181 GTGCTGGAACCGTACGCCGGAAGGGTAAGCGAATATGAACAGATCACCTCAATTCAG 240
 QY 241 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTCCTGAAGGGCGAGATCCGGAAG 300
 DB 241 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTCCTGAAGGGCGAGATCCGGAAG 300
 QY 301 ACAAAAGTGATGCTTTCACCGTCCAGGAGGGGGTACTACCTGAATACCAGTCAAAA 360
 DB 301 ACAAAAGTGATGCTTTCACCGTCCAGGAGGGGGTACTACCTGAATACCAGTCAAAA 360
 QY 361 CTGGAAGTTAAACAAATCTGATTAGCGGCAGATCTTAAAGGGGGATGCAAAAACAATTCAC 420
 DB 361 CTGGAAGTTAAACAAATCTGATTAGCGGCAGATCTTAAAGGGGGATGCAAAAACAATTCAC 420
 QY 421 GGTGTTATTACCGGTATCTTCGAAGCGCGTATGGCAAAATAACGCCGGACCACTTAACCCGTG 480
 DB 421 GGTGTTATTACCGGTATCTTCGAAGCGCGTATGGCAAAATAACGCCGGACCACTTAACCCGTG 480
 QY 481 ATGAACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 540
 DB 481 ATGAACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 540
 QY 541 GCGTGGTATCTACACCGTTTTCAGGTGATGGGGGATAAAGCGCTGGAGCATTTGCTGCGGGC 600
 DB 541 GCGTGGTATCTACACCGTTTTCAGGTGATGGGGGATAAAGCGCTGGAGCATTTGCTGCGGGC 600
 QY 601 AATATTGTAATGATGCTGAACATGCTGACAGGCACTAGTCCACAGAAAGGTCGCTAT 660
 DB 601 AATATTGTAATGATGCTGAACATGCTGACAGGCACTAGTCCACAGAAAGGTCGCTAT 660
 QY 661 GTGCTGGAACCGTACGCCCGGAAGGGTAAGCGAATATGAAGAGATCACCTCAATTCAG 720
 DB 661 GTGCTGGAACCGTACGCCCGGAAGGGTAAGCGAATATGAAGAGATCACCTCAATTCAG 720

QY 1801 GCGTTGTTGCGGCAAAATCGCTGACGCTACGCGTTAAATTAACATA 1848
 DB 1801 GCGTTGTTGCGGCAAAATCGCTGACGCTACGCGTTAAATTAACATA 1848

RESULT 2
 AAA97430
 ID AAA97430 standard; DNA; 4115 BP.

XX AAA97430;
 AC
 XX
 DT 29-JAN-2001 (first entry)

XX Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH) gene.

DE D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
 KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
 KW L-ascorbic acid biosynthesis; vitamin C; ds.

XX Gluconobacter oxydans.

XX WO200055329-A1.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-JP01608.

XX 17-MAR-1999; 95JP-0072810.

XX 06-AUG-1999; 95JP-0224679.

XX (FUJI) FUJISAWA PHARM CO LTD.

XX Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
 PI Yamashita M, Takata Y;

XX WPI; 2000-587530/55.

XX P-PSDB; AAB23172.

XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 PT acid production

XX Claim 9; Page 60-63; 72pp; Japanese.

XX The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorbose by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
 CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the Gluconobacter
 CC oxydans SLDH gene.

XX Sequence 4115 BP; 833 A; 1133 C; 1270 G; 879 T; 0 other;

Query Match 5.7%; Score 105; DB 21; Length 4115;
 Best Local Similarity 46.8%; Pred. No. 4.6e-24;
 Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;

QY 503 TCCACATCGGTTAGTCTTTTCATCGCGCACATCAGCGGTGATCTACACCGTTTCC 562

DB 622 TCCATTTGGTGTAGTAACTTTTTCGAGCCCATCAGCGGTCTACGTCGACGATTC 681

QY 563 AGTGATGGCGGATAAACGCTGAGACATGTCTCGGGCAATATTCGTAATGA---TGTGTG 619
 DB 682 TTGAACACGCTCCGACATGGGCGATTGTTGGTGTGGCCGTACGGGCACTGCGTTCAA 741
 QY 620 AACATGTCGTACAGCAGCTAGTCACAGAAAGTTCGCTATGTCTGGGAAACCGTCAGCC 679
 DB 742 AGAAAAAGCCGAGGAATTCAGGCCAGGACTGCTGTATTTCCTTGACCGAGAGCGGCTC 801
 QY 680 CGGAAGGGGTAAAGCAATATGAAGAGATCACCTCAATTCAGAGATGTGATACCGTGGCAGG 739
 DB 802 CGTCCGGCAAGAGACAGCGGTGCGGCTCATGGGCGCGCTGCTGCTACTATCTGCTTGC 861
 QY 740 CAGATTTACACCGCTGATGTGAAGGGGAGATCCGGAAGCAAAAGTGTGTTTCA 799
 DB 862 CCGATCCGGAAGCCGTGCTGAAGCATCTTGTGTGATCCCGCCATCGGTTTCCATGA 921
 QY 800 CGGTACCGCAAGCGGGGTACTACTCTGA---ATACAGGTACAAACTGGAAGTTAAACAATC 856
 DB 922 CGATCAGGAAGCGCGCTACAACATCAGGAGACACCGGTGCTGCTGCGATCTGGAGATG 981
 QY 857 CTGATTTAGCGGAGATCTTAAAGGGGATGCAAAACAAATTTACGGTGTATTATACCCGTA 916
 DB 982 CGGCAGTAAAGCCGACCTCAAGAACCCGGAAGACCGGCTACCGGTTTTCGGTTACGTG 1041
 QY 917 TCCTCGAAGCGGTATGCAAAATAACGCCG-----ACCACTAACCCCTGCTGAATTCGG 970
 DB 1042 TCGAGGCCCTGCTGCTGCTGGGATGCGCGTGTGAAGCATTTACGGTCTATGCTCCTGTG 1101
 QY 971 ATAAGTCCGCCATATGCTGAACGTTTCCATGATGGCTGCTGAGTTTCTCCAGCTAA 1030
 DB 1102 ATAACTCGGTCATACCGCAATGTGCCCGCAAGCCCTTCTCTGCTATGCGAAG---G 1158
 QY 1031 CTGGCAACAGGATGTCACTGCTGCTGAGTACAAATACCACTTCCCGCAATACCATGG 1090
 DB 1159 CCGCGATCCGAGTTGCGCAAGTGGATTGAGGAAACCGGACCTTCCCGAAGCAATG 1218
 QY 1091 TTGACCGCATTAAGCTCTCGCGGAGAGAACTTCCGGCAGCGGATCAAGGCTCAACCG 1150
 DB 1219 TTGATCGCATCACCGGACCGTTTTCGGCGGAAATCGCAAGAGCTCAACGCGGCCAGT 1278
 QY 1151 GTATTGCCGATAAAGCGCGGTAATGGGCAAAACCTTTATCCAGTGGGTGCTGGAAGATA 1210
 DB 1279 GCGTGTACGACCTGCGGCTGCTGGCGAGGATTTCCATCAGTGGGTGCTGGAAGACC 1338
 QY 1211 ATTTCGCTGATGCTCGGCACTGGAGAAGTTCGCTGCGCAACTGGGTGGGCTCGGTAA 1270
 DB 1339 AGTTTCGGATGCGCTCCGCCCTTGAAAAAGCCGCGTGCAGATGGTTCGGGAGCTGA 1398
 QY 1271 TCCCTATGAAGAGGGGAAGATTTCGATTTCTTAACCTTTACACACAGTTGCTGCTGGG 1330
 DB 1399 CGGACTGGGATACGTTCAAGATCCGAATGCTCAATGCAGGGCATGCTATGCTCTGCTTC 1458
 QY 1331 CAGGTACGTTAATCGGTCAAAAATATATCCAGAAAGCAATGACCGGATTTATCTATC 1390
 DB 1459 CAGGATTTCTGGTGGGCTATGAGATGTGGATGACGCCATTGAAGACAGCAACTCCTTG 1518
 QY 1391 AGATTGCCACCGCTACGTGACGGAAGATGCTATTCTCTGCTT---GGCGATAACGGTA 1447
 DB 1519 GCAATCTGAAGAACTATCTCAACAGGATGTATCCCGACCTGAAGGCGCCCTTCAGGCA 1578
 QY 1448 TCGATTTGCCAACCTACCGTGTGTTGTTACTCAAGCGTTTTTACCAATCACCATATTTCAG 1507
 DB 1579 TGACGCTCGAAGGCTATCGGGACAGCGTCTCAGCGGTTTCTCCAACAAGCGCATGCGG 1638
 QY 1508 ACACCAACCAAGCGCTGCTGGGATGTTTCTCGAAAAATTCGGCGGATGATTTGCCCCCA 1567
 DB 1639 ACCAGACGCTCGGATGTTGATGCGATGGTGTTCGAAGTTTCAGGTGTTCTGGACGAAA 1698
 QY 1568 CACTGG 1574
 DB 1699 CCGTGG 1705

```
RESULT 3
AAZ35672
ID AAZ35672 standard; DNA; 1458 BP.
XX
AC AAZ35672;
XX
DT 27-JAN-2000 (first entry)
XX
DE Gluconobacter suboxydans L-sorbose reductase nucleotide sequence.
XX
KW Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
KW mutant; ds.
XX
OS Gluconobacter suboxydans.
XX
FH Key Location/Qualifiers
FT CDS 1..1458
FT /tag=a
FT /product="L-sorbose reductase"
XX
PN AU9920390-A.
XX
PD 23-SEP-1999.
XX
PF 11-MAR-1999; 99AU-0020390.
XX
PR 13-MAR-1998; 98EP-0104546.
XX
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
XX
WPI; 1999-579276/49.
DR P-PSDB; AAY49913.
XX
PT Set of genetically modified mutants not containing L sorbose reductase
PS Example 4; Page 17-18; 33pp; English.
XX
CC The present invention describes a genetically engineered
CC microorganism derived from a microorganism belonging to the genus
CC Gluconobacter or Acetobacter which is characterised in that the
CC biological activity for reducing L-sorbose is substantially
CC nullified by gene recombination. The present sequence encodes
CC Gluconobacter suboxydans L-sorbose reductase.
CC N.B. This patent is equivalent to the basic; NO9901197 in week 199949.
XX
SQ Sequence 1458 BP; 329 A; 386 C; 434 G; 309 T; 0 other;

Query Match 5.5%; Score 101.8; DB 20; Length 1458;
Best Local Similarity 51.4%; Pred. No. 3e-23;
Matches 310; Conservative 0; Mismatches 287; Indels 6; Gaps 3;

QY 953 TAACCTGCTGAATTCGGATAACGTGCGCAATAATGCTGAACGTTTCCATGATGCGCTGG 1012
DB 548 TCACGATCATGCTCGGATAATCTCGGCATAACGGT-AATGTCGCCCGCAAGGCATTT 606
QY 1013 TTGAGTTTCTCCAGCTAACTGGCAACAGAGATGTATCGACTGGCTGAGTACAAATACCA 1072
DB 607 CTGGGATACGC--GAAGGCCCGGTGATCCGCAACTGGCCAAAGTGGATTGAAGAGACGCCGA 664
QY 1073 CTGCGCCGAATACCATGTTGACCGCATTAACGCTCGTCCGGCAGCAGAACTTCCGGCAC 1132
DB 665 CGFTCCCAATGGCATGGTTGATCGCATCATCGCCGACCGCTTCTGCTGACATTGCGGAAGA 724
QY 1133 GGATCAAGGCTAAACGGGTATTTCGGGATAAAGCGCGGTAAATGGCGAAACCTTTATCC 1192
DB 725 AGCTCAACCAAGCCAGTGGCTGCAGCAGACACCTGCGGCTGTTGACAGAGACTTTTCATC 784
QY 1193 AGTGGTCTGGAAGATAATTTCCGTGATGTCCGTCGCCCACTGGAGAGAGGTCGGTGCG 1252
DB 785 AGTGGTCTGGAAGACAGCTTTTGCTGATGGCGCGGCTCGCTGGAAAGAGCGCGGAGTGC 844
```

```
QY 1253 AACTGTGGCGTCGGTAATCCCCCTATGAGAGCGCGAAGATTCGCATTTCTTAACCTTTCAC 1312
DB 845 AGTTTCGTTGGGATGTGACGGACTGACGAGCATGTAAAAATCCGCATGCTGAATGCTGGTC 904
QY 1313 ACAGTTGCATCGCCTGGGAGGTACGTTAATCGTCAAAATATATATCCAGAAAGCACAA 1372
DB 905 ACATCATGCTCTGCTTCCCGGCTGTTCTGGCAGGATTTGAAAATGTCGATCATGCCCTTG 964
QY 1373 TGACCGGATTTTATCTATATCAGATTGCCGACCGCTACGTCGACGGAAGATGTCTATTCCTTGCT 1432
DB 965 CTGATCCCGATCTACCGCGTATCTCTCGAGAACTTCTGACAAAGAGCGTCATCCGACCC 1024
QY 1433 T---GGCGGATAACGTATGATTTGCCAACCTACCGTATGTTGTACTCAAGCGTTTAA 1489
DB 1025 TGAAGGCACCGCGGCGATGACGCTGGAAGGCTATCGGACAGCGTGATCAGCGGTTCT 1084
QY 1490 CCAATCCACATATTCAGGACACCAACCAACGCGCTCGGATGTTTCTCCAAATTC 1549
DB 1085 CGAATCCGGCCATGCGCGGATCAGACATTCGTTATTTCCGGGGACGGAGCTCGAAGATCC 1144
QY 1550 CGG 1552
DB 1145 AGG 1147

RESULT 4
AAZ32025
ID AAZ32025 standard; DNA; 9810 BP.
XX
AC AAZ32025;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human METH1 related EST AF018073.
XX
KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX
OS Homo sapiens.
XX
PN WO9937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
XX
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX
WPI; 1999-590684/50.
XX
PT New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
XX
PS Disclosure; Page 353-359; 457pp; English.
XX
CC AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
```


Db 857 CTGCTGGGAAGTCCAGGCTGACAAATGGTGAATGATGATGCTCTGCCATGGGAAGAGATGA 798
Qy 1289 AGATTTCGATCTTAACTCTTCACACAGTTGTCATCGCTGGGAGGTACGTTAAATCGGTC 1348
Db 797 AACTCGGATGCTTAATGGCAGCACTCTTTCTCGCTTATCTGCTTACCTCTCAGGAT 738
Qy 1349 AAAATATATCCAGAAAGCACAATGACCGATTTTATCTATCAGATTCGCGACCGCTACG 1408
Db 737 TCGCCCATATGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 678
Qy 1409 TCACGGAGATGCTATCTTCTTCTTGGGGATTAACGGTATCGATTTGCCAACCTACCGTG 1468
Db 677 TGCTGGATGAGAACGCGGACACTGCAAAATTAAGATGTCGATTTAACACAATATGCGG 618
Qy 1469 ATGTTGTACTCAAGGTTTTTACCAATCCACATATTCAGGACACCAACCAACGCTGCTG 1528
Db 617 ATAGTTAATTCGACGTTTTGCTAATCCGCGCTGAACATTAAGACCTGGCAATTCGCGA 558
Qy 1529 CGGATGTTTTTCGAAAATTCGG 1551
Db 557 TGGATGGCAGCAGAAATTAACG 535

RESULT 7
AAS78056
ID AAS78056 standard; cDNA; 2810 BP.
XX
AC AAS78056;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13860.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG13869.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 13860; 103pp; English.
XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2810 BP; 696 A; 699 C; 758 G; 657 T; 0 Other;
Query Match 4.0%; Score 73.4; DB 23; Length 2810;
Best Local Similarity 49.1%; Pred. No. 1.9e-13;
Matches 194; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
Qy 1165 GCGCCGGTAATGGGCGAAACCTTTATCCAGTGGTGGTGGTGAAGATAAATTCGTTGATGTC 1224
Db 2 GCGGCGGTTGCTGTGAACCTTTCCGCCAGTGGGTAATAAGATAAATTCGTTGCCGA 61
Qy 1225 CGTCCGGCACTGGAGAAGTGGTGTGCAACTGGTGGCGTGGTGAAGATAAATTCGTTGATGTC 1284
Db 62 CGTCCGGTAATGGGCGAAAGCGGAGCGGCACTGGTGAAGATAAATTCGTTGATGTC 121
Qy 1285 GCGAAGATTGCGATCTTTAACTCTTCACACAGTTCGATCGCTGGGAGGATACGTTAATC 1344
Db 122 ATGAAGTTGCGCATGCTCAACGGCAGTCAATTCCTGGCGTATCTGGGTATCTTGA 181
Qy 1345 GGTCAAAATATATCCAGAAAGCACAATGACCGGATTTTATCTATCAGATTCGCGACCGC 1404
Db 182 GGATATCAGCACATTAATGACTGTATGGAAGATGAACATTTATCGTTATCGCGGTATGC 241
Qy 1405 TACGTGACGGAAGATGTCATCTTCTTGGGCGATAACGGTATCGATTTGCCAACCTAC 1464
Db 242 TTGATGTTGCAGGAACAACGCGCGCGGTTGAAGTGCAGGCGGTTGATTGCAAGATTAC 301
Qy 1465 CGTGATGTTGTACTCAAGCGTTTTTACCAATCCACATATTCAGGACACCAACCAACCGTC 1524
Db 302 GCTAACCGGATTAATTCACGCTATAGCAACCGCGGTTACGTCATCGAACCTGGCAGATT 361
Qy 1525 GCTGCGGATGTTTCTCGAAATTTCCGGCGATGAT 1559
Db 362 GCGATGGATGGTAGCCAGAAATTCGCACAGCGGAT 396

RESULT 8
AAH65091
ID AAH65091 standard; DNA; 1509 BP.
XX
AC AAH65091;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 126.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Db	121513	GGCGCCCCCGTACGAGGAGTTGGCGTGCAGTCTCTCCGACGCTGGAGCCTTATGAA	121554
QY	1282	GAGCGGAAGATTCCGATTCCTTAACTCTTCACACAGTTGTCATCGCCTGGGAGGTACGTTA	1341
Db	121553	TTAATGAAGCTGGCGCTGCTCAACGGCTCCACACAGGACTTTGCTACTTCGGCCACTTG	121494
QY	1342	ATCGGTCAAAATATATCCACGAAAGCAACAATGACCGATTTTATCTATCAGATTGCCGAC	1401
Db	121493	GCTGGCCACCACATGGTGCACGAGCTATGGCGGATACCGGCTTCAGAGATTTCTCTCTG	121434
QY	1402	CGTTACGTGACGGAAGATGTCATTCCTTGCTT---GGGCGATAACGGTATCGATTTGCCA	1458
Db	121433	GCTTACATGGGCGGAGGCCACCCCTACCTCAAGAACTCCAGGTGTCGATCTAGAT	121374
QY	1459	ACCTACCGTGATGTTGTACTCAAGGGTTTTACCAATCCACATATTCAGGACACCAACCAA	1518
Db	121373	GCTTATCGACGCCAACTCATPCTGGCGGATTCGGCAACGCCGAGTCAAGACACCGTACCG	121314

CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

CC Sequence 684 BP; 152 A; 212 C; 185 G; 135 T; 0 other;

Query Match 3.3%; Score 61.2; DB 21; Length 684;
 Best Local Similarity 54.4%; Pred. No. 1.1e-09;
 Matches 123; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 1034 GCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTTCCCGAATACCATGTTG 1093

Db 8 GCACCCCGAGATCGGAATGGATTCGCCGACAGGGCGCTTCCCAACCCCATGGTCG 67

Qy 1094 ACCGCATTACGCCCTCGTCCGCGACGACAGAACTTCGCGCACGGATCAAGGCTCAACACGGGTA 1153

Db 68 ACCGTATCACCCCTCAGACATCCGCCACCGGATAAACGGCACTCGCAGACAACTTGCCA 127

Qy 1154 TTCCCGATAAGCGCGGTAAATGGCGGAACCTTTATCCAGTGGTGTGGAAGATAATT 1213

Db 128 TCGAGGACTCGTGGCCCGCTCGTCACAGAGCCCTTTATGCACTGGGTAAATGAGGATCAGT 187

Qy 1214 TCCGTGTATGTCGTCGCGCACTCGAGAGGTCGCTGTCGAACCTGGT 1259

Db 188 TCTCGATGGCGGCCACCATTCCAGAGAGGTGTGTCGCCAGTGGT 233

RESULT 13

AA582343
 ID AAS82343 standard; cDNA; 2193 BP.

AC AAS82343;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #18147.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG18156.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID NO 18147; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcc_sequences.

CC Sequence 2193 BP; 494 A; 592 C; 603 G; 504 T; 0 other;

Query Match 2.3%; Score 43.4; DB 23; Length 2193;

Best Local Similarity 66.7%; Pred. No. 0.0025;

Matches 62; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1168 CCGGTAATGGCGCAACCTTTATCCAGTGGTGTGGAAGATAATTTCCGTGATCCGT 1227

Db 799 CAGATGACTGTGCAACCGTTATCCAGTGGTGTGGAAGATAACTTCGCTGGCGGT 858

Qy 1228 CCGGCACTGGAGAGGTGGTGTGCAACTGGTG 1260

Db 859 CTGCGCTGGGAAGTCGCAGGTGTACAAATGGTG 891

RESULT 14

AAH48024

ID AAH48024 standard; DNA; 14041 BP.

AC AAH48024;

DT 18-SEP-2001 (first entry)

DE Internal control B19c #1.

XX Internal control; ss.

OS Parvovirus.

XX WO200146463-A2.

XX 28-JUN-2001.

XX 20-DEC-2000; 2000WO-EPI2996.

XX 22-DEC-1999; 99AT-0002170.

XX (BAXT) BAXTER AG.

XX Zimmermann-K, Turecek P, Schwarz H, Rieger M;

XX WPI; 2001-408658/43.

PT Internal standards useful for nucleic acid amplification assays,
 PT comprises a synthetic nucleic acid made by non-recombinant techniques
 XX -
 XX Example 1; Page 22-27; 30pp; English.

XX The present invention relates to methods for the preparation and use of
CC internal controls for nucleic acid amplification assays. The internal
CC controls comprise a synthetic nucleic acid made by non-recombinant
CC techniques. The internal controls are useful for detecting nucleic acids
CC in a sample such as blood, spinal fluid, semen, saliva, tears, cell
CC culture fluid, recombinant cells, animal tissue or plant tissue, by a
CC quantitative PCR assay, by adding the internal control to the sample,
CC amplifying the nucleic acids in the sample and detecting the amplified
CC products. The internal controls help in performing the nucleic acid
CC amplification assay quickly and inexpensively without sacrificing assay
CC specificity or sensitivity. The present sequence is one such internal
CC control, which was used in an example to illustrate the present
CC invention.
CC Note: the present sequence is the SEQ ID 8 shown in the sequence listing.
CC This sequence differs from the SEQ ID 8 shown on page 13 of the
CC disclosure (see AAH48038).
XX
SQ Sequence 14041 BP; 2151 A; 1354 C; 466 G; 2218 T; 7852 other;
Query Match 2.3%; Score 42.4; DB 22; Length 14041;
Best Local Similarity 18.8%; Pred. No. 0.017;
Matches 72; Conservative 132; Mismatches 177; Indels 3; Gaps 1;
QY 1285 GCGAAGATTCGCAATCTTAACCTTCACACAGTTGCATCCCTGGCAGTACGTTAATC 1344
DB 2734 SSSCHGNCDSWTHNANTRMYSHRTTMANDATVRYCNCSTASARSCADGNCDSYNTH 2793
QY 1345 GGTCAAAATATATCACGAAGACCAATGACCGATTTCATCATCAGATTGCGGACCGC 1404
DB 2794 SRSSCHASRAMGSYNTHSSRWGBCUAGBNBGRGMANYTHANTCATNTHGNCDSYN 2853
QY 1405 TAGCTGACGGAAGATGTCATCTCTGTCGGCGATAACGGTATCGATTGCGCAACCTAC 1464
DB 2854 THSDSAYTAKSACATMATCAYNTHSYNTHSRRTSRADYRVDBYTHSYNTHSRYSYTN 2913
QY 1465 CGTGATGTTGCTACTCAAGCGTTTACCAATCCACATATTCAGCACACCAACCGGTC 1524
DB 2914 ADD---TNTHSANTCATNSNTNRRDBYTHRSNCRTNSWHCHCANBWNATDNYASARST 2970
QY 1525 GCTGCGGATGTTTCTCGAAATTCGGCGGATGATTGCCGCCACACTGGGAGAGTGCTAC 1584
DB 2971 MCNSMNGNSVRCATNANDTRACNRCSSNTHCASHBGNCRARATNSTANDARDSNTHR 3030
QY 1585 CACGAGGCGTTCGCCGAATGCCACCGCATGTTACCTGCACTGTTTACGTTATTCATG 1644
DB 3031 RARTSRNSGYTHRCSSNACCRDANCWTHNVTNNSRBYWSTDTARGSCANDSTRASA 3090
QY 1645 GAGCAGTGGCATCACGGCAAACTG 1668
DB 3091 NDCMRHNSVMDASCRRNGTSTS 3114
RESULT 15
AAS92760
ID AAS92760 standard; cDNA; 2238 BP.
XX
AC AAS92760;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28564.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX P-PSDB; ABC28573.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 28564; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2238 BP; 600 A; 491 C; 577 G; 570 T; 0 other;
Query Match 2.1%; Score 38.4; DB 23; Length 2238;
Best Local Similarity 45.6%; Pred. No. 0.13;
Matches 135; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 1256 TGTGCGCGTGGTAATCCCTATGAAGAGCGGAGATTCGCATCTTAACCTTTCACACA 1315
DB 2 TGTGTAATGATGTCCTGCCATGGGAAGAGATGAACCTGGGATGCTTATGGCAGCCACT 61
QY 1316 GTTGCAATCGCTGGGAGGTACGTTAATCGGTCAAAAATATATCCACGAAGCAATGA 1375
DB 62 CTTTCTCGCTTATCTGGGTTACCTCTCAGGATTCGCCCATATCATGATTCATCAGC 121
QY 1376 CCGATTTTATCATCAGATTGGCGACCGCTACGTGACGGAAGATGTCATCTTGGTGG 1435
DB 122 ATCGCGCATTTTCGCCATGCCCGCAGAACATTAATGCTGGATGAGCAAGCCGACACTGC 181
QY 1436 GCGATAACGGTATCGATTTGCCAACCTACCGTGATGTTGTACTCAAGCGTTTTTACCAATC 1495
DB 182 AAATTAAGATGTCGATTTAACAACATATATGCGGATGAAGTTAATGACGCTTTTGCCTAATC 241
QY 1496 CACATATTCCAGGACACCAACCGCTCGCTCGGATGTTTCTCGAAATTCG 1551
DB 242 CGGCGCTGAACATAAGACCTGGCAAATCCGATGCGATGCGACGCAAGAAATACCG 297

Search completed: March 21, 2003, 20:48:30
Job time : 805.996 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 09:44:37 ; Search time 67.0095 Seconds
(without alignments)
8457.593 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgaacgaacaattacatg.....tctacgctgtaataactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44.4	2.4	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	33.2	1.8	4797	4 US-09-643-597-134	Sequence 134, App
C 3	32.4	1.8	1560	4 US-09-453-702B-264	Sequence 264, App
C 4	32.4	1.8	2230	4 US-09-342-647-1	Sequence 1, Appli
C 5	32.4	1.8	3271	4 US-09-770-170-3	Sequence 3, Appli
C 6	32.4	1.8	61663	4 US-09-453-702B-62	Sequence 62, Appl
C 7	32	1.7	1407	1 US-08-459-287-1	Sequence 1, Appli
C 8	31	1.7	2796	2 US-08-937-931-3	Sequence 3, Appli
C 9	31	1.7	2796	4 US-09-285-502-3	Sequence 3, Appli
C 10	31	1.7	2796	4 US-09-709-126-3	Sequence 3, Appli
C 11	31	1.7	2796	4 US-09-871-385A-3	Sequence 3, Appli
C 12	31	1.7	3349	2 US-08-920-234-1	Sequence 1, Appli
C 13	31	1.7	3410	4 US-09-527-154-3	Sequence 3, Appli
C 14	30.8	1.7	1339	1 US-07-936-163-2	Sequence 2, Appli
C 15	30.6	1.7	699	4 US-08-998-416-591	Sequence 591, App
C 16	30.6	1.7	3200	1 US-08-444-405-1	Sequence 1, Appli
C 17	30.6	1.7	3200	1 US-08-384-850-1	Sequence 1, Appli
C 18	30.4	1.6	541	4 US-09-221-017B-68	Sequence 68, Appl
C 19	30.2	1.6	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 20	30	1.6	1240	1 US-08-103-998-1	Sequence 1, Appli
C 21	30	1.6	1661	1 US-08-518-474-1	Sequence 1, Appli
C 22	29.8	1.6	216	1 US-07-663-413-28	Sequence 28, Appl
C 23	29.8	1.6	216	1 US-08-055-530-28	Sequence 28, Appl
C 24	29.8	1.6	216	1 US-08-247-475-47	Sequence 47, Appl
C 25	29.8	1.6	216	1 US-08-479-650-47	Sequence 47, Appl
C 26	29.8	1.6	216	1 US-08-191-866D-75	Sequence 75, Appl
C 27	29.8	1.6	216	1 US-08-674-169-47	Sequence 47, Appl

C 28	29.8	1.6	216	2 US-08-185-949B-75	Sequence 75, Appl
C 29	29.8	1.6	1172	2 US-08-070-301-21	Sequence 21, Appl
C 30	29.8	1.6	1400	2 US-08-305-764C-57	Sequence 57, Appl
C 31	29.8	1.6	1400	2 US-08-305-764C-59	Sequence 59, Appl
C 32	29.8	1.6	2095	2 US-08-305-764C-55	Sequence 55, Appl
C 33	29.8	1.6	3226	2 US-08-070-301-10	Sequence 10, Appl
C 34	29.8	1.6	4403765	4 US-09-103-840A-2	Sequence 2, Appli
C 35	29.8	1.6	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 36	29.6	1.6	1370	4 US-09-221-017B-716	Sequence 716, App
C 37	29.6	1.6	1436	4 US-08-961-527-365	Sequence 365, App
C 38	29.6	1.6	4483	4 US-08-961-527-363	Sequence 363, App
C 39	29.6	1.6	11717	1 US-08-801-263A-4	Sequence 4, Appli
C 40	29.6	1.6	11717	3 US-09-102-248-4	Sequence 4, Appli
C 41	29.6	1.6	32768	4 US-08-961-527-71	Sequence 71, Appl
C 42	29.4	1.6	1550	4 US-08-858-207A-74	Sequence 74, Appl
C 43	29.4	1.6	2625	4 US-09-030-335-8	Sequence 8, Appli
C 44	29.4	1.6	8937	2 US-08-449-933-1	Sequence 1, Appli
C 45	29.4	1.6	8937	4 US-07-966-049A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match 2.4% Score 44.4; DB 1; Length 7218;

	Query Match	1.8%	Score 32.4	DB 4	Length 3271
	Best Local Similarity	53.1%	Pred. Nsm. 2.2		
	Matches 69	Conservative 0	Mismatches 61	Indels 0	
y	951	ACTAACCTCTCTCAATTGGCATACGTGCGCCATAATGGTGAACGTTTCCATCATGCGG			
b	2290	AGTCATACGCTTTATGGTGAAGTCTCTGGTCAATTATATGTTCTCGCTCTTGATGATG			
y	1011	GGTTGAGTTTCTCCAGCTAACTGGCAACAGGATGTCGACGTGGCTGAGTACAAAT			

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,931
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2796 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-285-502-3

Query Match 1.7%; Score 31; DB 4; Length 2796;
Best Local Similarity 45.9%; Pred. No. 6.1;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1264 TCGGTAATCCCTATGAGAGCGGAGATTCCGATCTTAACTCTTCACACAGTTGCGATC 1323
DB 913 TCCGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAAGTATTTTTC 854
QY 1324 GCGTGGCAGGTACGTTAATCGGTCAAAATATATCCAGGAAAGCAATGACCGATTTT 1383
DB 853 AGCTGAATTTGACGTTTTTCTCAGAAAGTTCTGGACCATAGCAGCATGTTCTCTTG 794
QY 1384 ATCTATCAGATTCGCGACCGCTACGTCGAGGAGATGTCATTCCTTGCTGGCGGATAC 1443
DB 793 AGGTATCTGTGTTACTCTCTACACCATGTCATCTGGTATTTCTTCATTTTCAAATAC 734
QY 1444 GGTATCGATTTGCCAACCTACCGTGATGTTGACTCAAGCGTTTACCAT 1494
DB 733 TGAATGATCTGCACAGCCCCCTGAGGACCGTATTTATGGGATAGTTAAT 683

RESULT 10
US-09-709-126-3/C
Sequence 3, Application US/09709126
Patent No. 6319704
GENERAL INFORMATION:
APPLICANT: Rubin, Gerald M.
Pan, Duojia
Rooke, Jenny
Yavari, Reza
Xu, Tian

TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/709,126
FILING DATE: 08-NO. 6319704-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/285,502
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2796 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-709-126-3

Query Match 1.7%; Score 31; DB 4; Length 2796;
Best Local Similarity 45.9%; Pred. No. 6.1;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 1264 TCGGTAATCCCTATGAGAGCGGAGATTCCGATCTTAACTCTTCACACAGTTGCGATC 1323
DB 913 TCCGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAAGTATTTTTC 854
QY 1324 GCGTGGCAGGTACGTTAATCGGTCAAAATATATCCAGGAAAGCAATGACCGATTTT 1383
DB 853 AGCTGAATTTGACGTTTTTCTCAGAAAGTTCTGGACCATAGCAGCATGTTCTCTTG 794
QY 1384 ATCTATCAGATTCGCGACCGCTACGTCGAGGAGATGTCATTCCTTGCTGGCGGATAC 1443
DB 793 AGGTATCTGTGTTACTCTCTACACCATGTCATCTGGTATTTCTTCATTTTCAAATAC 734
QY 1444 GGTATCGATTTGCCAACCTACCGTGATGTTGACTCAAGCGTTTACCAT 1494
DB 733 TGAATGATCTGCACAGCCCCCTGAGGACCGTATTTATGGGATAGTTAAT 683

RESULT 11
US-09-871-385A-3/C
Sequence 3, Application US/09871385A
Patent No. 6399350
GENERAL INFORMATION:
APPLICANT: Rubin, Gerald M.
Pan, Duojia
Rooke, Jenny
Yavari, Reza
Xu, Tian

TITLE OF INVENTION: KUZ: A No. 6399350el Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/871,385A
FILING DATE: 31-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/709,126
FILING DATE: 08-NO. 6399350-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:05:17 ; Search time 138.147 Seconds
(without alignments)
10373.549 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgaacgaacaattacatg.....tctacgcgttaataactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1848	100.0	1848	9	US-09-802-208B-1
2	68.8	3.7	1509	9	Sequence 1, Appli
3	36	1.9	653	9	Sequence 126, App
4	33.2	1.8	671	9	Sequence 402, App
5	33.2	1.8	4797	10	Sequence 346, App
6	33.2	1.8	4797	10	Sequence 134, App
7	33.2	1.8	4797	10	Sequence 134, App
8	33.2	1.8	4797	10	Sequence 134, App
9	32.8	1.8	596	9	Sequence 310, App
10	32.8	1.8	615	10	Sequence 4050, Ap
11	32.8	1.8	919	9	Sequence 258, App
12	32.4	1.8	1560	9	Sequence 264, App
13	32.4	1.8	2436	10	Sequence 7778, Ap
14	32.4	1.8	61663	9	Sequence 62, Appl
15	32	1.7	2139	10	Sequence 9961, Ap
16	31.6	1.7	1200	10	Sequence 3, Appli
17	31.6	1.7	1833	10	Sequence 18, Appl
18	31.6	1.7	1833	10	Sequence 21, Appl
19	31.4	1.7	2176	9	Sequence 8, Appli

20	31.4	1.7	2265	9	US-09-843-250-1	Sequence 1, Appli
21	31.4	1.7	2265	9	US-09-843-250-25	Sequence 25, Appl
22	31.4	1.7	2265	9	US-09-843-250-27	Sequence 27, Appl
23	31.4	1.7	2265	9	US-09-843-250-28	Sequence 28, Appl
24	31.4	1.7	2265	9	US-09-843-250-29	Sequence 29, Appl
25	31.4	1.7	2265	9	US-09-843-250-30	Sequence 30, Appl
26	31.4	1.7	2265	9	US-09-843-250-31	Sequence 31, Appl
27	31.4	1.7	2265	9	US-09-843-250-56	Sequence 56, Appl
28	31.4	1.7	2265	9	US-09-843-250-57	Sequence 57, Appl
29	31.4	1.7	2294	9	US-09-843-250-6	Sequence 6, Appli
30	31.4	1.7	2515	9	US-09-843-250-4	Sequence 4, Appli
31	31.4	1.7	4355	9	US-09-843-250-7	Sequence 7, Appli
32	31.4	1.7	4744	10	US-09-775-938A-32	Sequence 32, Appl
33	31.4	1.7	9706	9	US-09-843-250-5	Sequence 5, Appli
34	31.4	1.7	9841	9	US-09-843-250-3	Sequence 9, Appli
35	31.4	1.7	14462	9	US-09-843-250-9	Sequence 9, Appli
36	31.2	1.7	3262	10	US-09-764-869-2430	Sequence 2430, Ap
37	31.2	1.7	3262	10	US-09-764-869-2433	Sequence 2433, Ap
38	31	1.7	341	10	US-09-770-791-841	Sequence 841, App
39	31	1.7	341	10	US-09-924-035A-866	Sequence 866, App
40	31	1.7	783	9	US-09-938-842A-1409	Sequence 1409, App
41	31	1.7	1800	9	US-10-101-464A-849	Sequence 849, App
42	31	1.7	2594	10	US-09-070-927A-187	Sequence 187, App
43	31	1.7	2796	10	US-09-871-388-3	Sequence 3, Appli
44	30.8	1.7	521	10	US-09-815-242-2460	Sequence 2460, Ap
45	30.8	1.7	741	10	US-09-974-300-4674	Sequence 4674, Ap

ALIGNMENTS

RESULT 1
US-09-802-208B-1
; Sequence 1, Application US/09802208B
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parrott, Wayne
; APPLICANT: LaFayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Arabitol or Ribitol As Positive Selectable Markers
; FILE REFERENCE: UGA-855R
; CURRENT APPLICATION NUMBER: US/09/802.208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-802-208B-1

Query Match	100.0%	Score	1848	DB	9	Length	1848	
Best Local Similarity	100.0%	Pred. No.	0					
Matches	1848	Conservative	0	Mismatches	0	Indels	0	Gaps
0								
Qy	1	ATGAACGAACAATTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG	60					
Db	1	ATGAACGAACAATTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG	60					
Qy	61	CGGTGGTATCTACACCGTTTCAGGTGATGGCGGATAAACGCTGGAGCATTCGTCGGGCG	120					
Db	61	CGGTGGTATCTACACCGTTTCAGGTGATGGCGGATAAACGCTGGAGCATTCGTCGGGCG	120					
Qy	121	AATATTCGTAATGATGCTGAACATGCTGACAGGCACCTCAGTGCACAGAAAGGTCGCTAT	180					
Db	121	AATATTCGTAATGATGCTGAACATGCTGACAGGCACCTCAGTGCACAGAAAGGTCGCTAT	180					
Qy	181	GTGCTGGAACCGTCAGCCCGAAGGGTAAAGCGAATATGAAGAGATCACCTCAATTTCAG	240					
Db	181	GTGCTGGAACCGTCAGCCCGAAGGGTAAAGCGAATATGAAGAGATCACCTCAATTTCAG	240					
Qy	241	AAGTTGATACCGTGGCAGGCAGATTTACACCGCTGATTGCTGAAGGGGCAGATCCGGAAG	300					
Db	241	AAGTTGATACCGTGGCAGGCAGATTTACACCGCTGATTGCTGAAGGGGCAGATCCGGAAG	300					

QY 1563 CCCACACTCGGAGAGTGCTTACCAGGCGGTTCCCGCAATGCCACCCCATGTTAC 1622
DB 247 ...MTCSHMTHSHSHS.TMMMYCC...CY.TYBTMM...A.H.HSAM.S.SSS.SN.188
QY 1623 TGCACGTGTTTACGTATTATCGAGCAGTGCGCATCACGGCAAACTGCCCTATCA 1676
DB 187 ..S.SBST.H.HSSTWYTNBSKM.T.AMYM.CSNHSHSHSHS.KYHSTTA 134

RESULT 5

US-09-735-705-134

; Sequence 134, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy R.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wang, AiJun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C14
; CURRENT APPLICATION NUMBER: US/09/735,705
; CURRENT FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = A,T,C or G

US-09-735-705-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1074 TTGCCGGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGGGCTCACACACCTGTAGCATTACTGGTAG 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCCCTGAGTGGCAGGAGGAGTACATTGTTTTCG 3279
QY 1194 GTGGGTCGTGGAAGATAATTTCCGTGATGT 1223
DB 3280 TGGGGAGGTCTAATCTAGATATCGACTTGT 3309

RESULT 6

US-09-850-716A-134

; Sequence 134, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07

; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = A,T,C or G
US-09-850-716A-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1074 TTGCCGGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGGGCTCACACACCTGTAGCATTACTGGTAG 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCCCTGAGTGGCAGGAGGAGTACATTGTTTTCG 3279
QY 1194 GTGGGTCGTGGAAGATAATTTCCGTGATGT 1223
DB 3280 TGGGGAGGTCTAATCTAGATATCGACTTGT 3309

RESULT 7

US-09-880-107-3939

; Sequence 3939, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 248199
; NAME/KEY: unsure
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3939

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1074 TTGCCGGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGGGCTCACACACCTGTAGCATTACTGGTAG 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCCCTGAGTGGCAGGAGGAGTACATTGTTTTCG 3279
QY 1194 GTGGGTCGTGGAAGATAATTTCCGTGATGT 1223

Db 3280 TGGGAGGCTTAATCTAGATATCGACTTGT 3309

RESULT 8

US-09-897-778-134
; Sequence 134, Application US/09897778
; Patent No. US20020147143A1

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.455C16

CURRENT APPLICATION NUMBER: US/09/897,778

CURRENT FILING DATE: 2001-06-28

NUMBER OF SEQ ID NOS: 467

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 134

LENGTH: 4797

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: 135, 501, 4421, 4467, 4468, 4698

OTHER INFORMATION: n = A,T,C or G

US-09-897-778-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;

Best Local Similarity 51.3%; Pred. No. 5;

Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1074 TTGCCCGAATACCATGTTGACCGATTACCGCTCGTCGGCAGCAGAACTTCGGGCACG 1133

Db 3160 TGGCTGGAAGATCTCGCGGTGGGCTTGCGGCTCACACACCTGTAGCACTTACTGGTAG 3219

Qy 1134 GATCAAGCTCAACGGGTATTGGCGATAAAGCCGGTAATGGCGAAACCTTTATCCA 1193

Db 3220 GACCAAGCATCTTGGGGGGTGGCGCTGAGTGGCGGGACAGGAGTCACTTTGTTTCG 3279

Qy 1194 GTGGCTGTGGAAGATAATTCCTGTGATG 1223

Db 3280 TGGGAGGCTTAATCTAGATATCGACTTGT 3309

RESULT 9

US-10-184-644-310/c

; Sequence 310, Application US/10184644

; Publication No. US20030044930A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: ACIDS ENCODING THE SAME

CURRENT APPLICATION NUMBER: US/10/184,644

CURRENT FILING DATE: 2002-06-28

Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-184-644-310

Query Match

Best Local Similarity 17.1%; Score 32.8; DB 9; Length 596;

Matches 68; Conservative 108; Mismatches 221; Indels 0; Gaps 0;

Qy 1284 GCGAAGATTCCGATCTTAACCTTCACACAGTTCGCTCGGCGAGGTACCTTAAT 1343

Db 471 SSCSNATANTCSSAASS.SNABAS.CSSBASS.SNATAST.SSAASHSNATASTCSSAA 412

Qy 1344 CGGTCAAAATATATCCACGAAGACAAATGACCGATTTTATCTATCAGATTGCCGACG 1403

Db 411 SS.SNATASBCSSAASS.SNATASTCSSBASS.SNATANTCSSAASS.SNATANTCSS.A 352

Qy 1404 CTACGTGACGGAAGATGTCATTCCTTGTGGCGGATACACGGTATCGATTTGCCAACCTA 1463

Db 351 SS.SNABAS.CSSBASS.SNATACTCSSAASHSNATASTCSSAASS.SNATASTCSSAA 292

Qy 1464 CCGTGATCTTGTACTCAAGCGTTTACCAATCCACATATTCAGGACACCAACGCGT 1523

Db 291 SS.SNATASTCSSBASSSNATASTCSSAASS.SNATACTCNSAAYS.SNATASTCSSAA 232

Qy 1524 CGCTCGGATGTTCTCGAAATTCGCGGATGATTCGCCACACACGCGAGAGTGCTA 1583

Db 231 SS.SNATASTYSSBASS.SNATASTCSSAASS.SNATASTCSSAASHSNATASTCSS.A 172

Qy 1584 CCAGCGAGCGTTCGCCCGAATGCCACGCCCATGTTTACCTGCACGTGTTTACGTATTCAT 1643

Db 171 SS.SNATASTYSSBASS.SNATASTCSSABSSCSNABASTCSS.ASS.SNATASTCSSAA 112

Qy 1644 GGAGCGTGGCATCAGGCAAACTCCCTATGATAT 1680

Db 111 SS.SNAT.S.CSSTAS..SNATAS..CSSAAD..SNAB 75

RESULT 10

US-09-815-242-4050/c

; Sequence 4050, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: Prokaryotes

FILE REFERENCE: ELITRA.011a

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

Prior Application Number: 60/191,078

Prior Filing Date: 2000-03-21

Prior Application Number: 60/206,848

Prior Filing Date: 2000-05-23

Prior Application Number: 60/207,727

Prior Filing Date: 2000-05-26

Prior Application Number: 60/242,578

Prior Filing Date: 2000-10-23

Prior Application Number: 60/253,625

Prior Filing Date: 2000-11-27

Prior Application Number: 60/257,931

Prior Filing Date: 2000-12-22

Prior Application Number: 60/269,308

Prior Filing Date: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

Query Match 1.8%; Score 32.4; DB 9; Length 1560;
Best Local Similarity 62.2%; Pred. No. 4.4;

Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 765 AGGGCGAGTCCGAGACAAAGTATTGCTTTACCGTCCAGGAGGGGTACTACCT 824
 Db 225 AGGCACAGTCCGTAGACAGAGTCTTCCCTTACCGTCCAGGAGGGGTACTACCT 166

Qy 825 GAATACACAGTCCAGAACTGGAA 846

Db 165 CGGATCCATCGTCAATCCGTAA 144

RESULT 13

US-09-815-242-7778/c

; Sequence 7778, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7778

; LENGTH: 2436

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(2436)

US-09-815-242-7778

Query Match 1.8%; Score 32.4; DB 10; Length 2436;

Best Local Similarity 53.1%; Pred. NO. 5.9;

Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 553 CACCGTTTTCAGGTGATGGCGGATAAAGCTGGAGCATTTGTCGGGCAATATTCGTAAT 512

Db 1857 CGCCTTCTGCGCGATACGGGGGATAGCTCTCGACCTTCTTGGCGGACGATCATCAT 1798

Qy 613 GATGCTGAACATGCTGATAGGCACTCAGTCACAGAAAGTGCCTATGTGCTGGAACCC 672

Db 1797 GTCGCGGATTCGTGCGACCAACCACCATGTCGCGAGGTGCTCAGCTCGCGGGCTC 1738

Qy 673 GTCAGCCCGG 682

Db 1737 GTCGTCGGG 1728

RESULT 14

US-10-114-170-62/c

; Sequence 62, Application US/10114170

; Publication No. US20030023075A1

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Nicole T.

; Perna, Nicole T.

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/114,170

; FILING DATE: 01-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/453,702

; FILING DATE: 03-DEC-1999

; APPLICATION NUMBER: 60/110,955

; FILING DATE: 04-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.95017

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 251-5000

; TELEFAX: (608) 251-9166

; INFORMATION FOR SEQ ID NO: 62:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 61663

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-114-170-62

Query Match 1.8%; Score 32.4; DB 9; Length 61663;

Best Local Similarity 62.2%; Pred. NO. 49;

Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 285 AGGGCGAGTCCGAGACAAAGTATGCTTTACCGTCCAGGAGGGGTACTACCT 344

Db 38189 AGGCACAGTCCGTAGACAGAGTCTTCCCTTTACCGTCCAGGAGGGGTACTACCT 38130

Qy 345 GAATACACAGTCCAGAACTGGAA 366

Db 38129 CGGATCCATCGTCATTCGGTAA 38108

RESULT 15

US-09-815-242-9961

; Sequence 9961, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:18:35 ; Search time 2836.31 Seconds
(without alignments)
10552.182 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgaacgacaattacatg.....tctacgcgttaactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_estchum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	90.8	4.9	801	17	BH395246 AG-ND-144
C 2	88.2	4.8	761	17	A2139030 SP.0178.B
C 3	81.6	4.4	506	12	BG608411 296895 MA
C 4	62.4	3.4	799	17	CNS01QW5
C 5	39.2	2.1	244	17	A2578838 28a07 Sho
C 6	37.8	2.0	391	17	AQ758091 HS_5552.A

7	37.6	2.0	632	13	BG922047
C 8	37.2	2.0	624	17	A2115205
C 9	36.6	2.0	787	17	CNS01QW7
C 10	36.4	2.0	550	12	BG268636
C 11	36.4	2.0	581	10	AW165487
C 12	36.4	2.0	619	10	AW171987
C 13	36.4	2.0	771	17	AQ500296
C 14	36.4	2.0	791	11	AY110366
C 15	36	1.9	533	17	A2163357
C 16	35.8	1.9	684	9	AA570826
C 17	35.6	1.9	404	12	BG724574
C 18	35.6	1.9	415	13	BM321987
C 19	35.6	1.9	419	9	AI757652
C 20	35.6	1.9	454	13	BM321994
C 21	35.6	1.9	483	12	BG561744
C 22	35.6	1.9	484	12	BG561164
C 23	35.6	1.9	485	10	BE028322
C 24	35.6	1.9	817	12	BG418689
C 25	35.4	1.9	334	12	BF420883
C 26	35.2	1.9	423	12	BE848253
C 27	35.2	1.9	677	17	A2568609
C 28	35.2	1.9	687	17	A2568679
C 29	35.2	1.9	716	14	BQ482891
C 30	35.2	1.9	975	17	CNS01IP2
C 31	35	1.9	486	10	AV634532
C 32	35	1.9	515	14	BQ809588
C 33	35	1.9	527	17	A2219883
C 34	35	1.9	718	14	BQ822949
C 35	35	1.9	782	9	AI068997
C 36	35	1.9	891	17	CNS03EWK
C 37	34.8	1.9	121	17	A2579022
C 38	34.8	1.9	488	14	BQ163971
C 39	34.8	1.9	535	14	BQ163790
C 40	34.8	1.9	541	13	BM660708
C 41	34.8	1.9	700	9	AL628755
C 42	34.8	1.9	1201	17	CNS01652
C 43	34.6	1.9	481	13	BM370033
C 44	34.6	1.9	649	10	BE049425
C 45	34.6	1.9	1101	17	CNS00LXJ

ALIGNMENTS

RESULT 1
BH395246/c
LOCUS
DEFINITION
AG-ND-144F15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-144F15
801 bp DNA linear GSS 11-DEC-2001
BH395246
DNA sequence.
BH395246
GI:17341387
GSS
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
1 (bases 1 to 801)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-144F15.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 506)
AUTHORS Fahrenkrug, S.C., Erking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith RPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980504.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACATGATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 1 row: A column: 22
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..506
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1PIC"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 130 a 118 c 138 g 120 t
ORIGIN

Query Match 4.4%; Score 81.6; DB 12; Length 506;
Best Local Similarity 49.1%; Pred. No. 7.4e-14;
Matches 216; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
QY 1112 CGGCAGCAGAACTTCGGCAGCGATCAAGGCTCAACAGGGTATTCGCCGATAAGCCCGG 1171
DB 1 CTGCAACGACAACTACTGGTGGAAATCAGCCAGCATCTGGGGGTGAATGATCCTCGCGCA 60
QY 1172 TAATGGCGCAACCTTTATCCAGTGGTCTGGAAGATAATTTCCGTGATGTCGCGCG 1231
DB 61 TTAGCTGCGAAGCTTTATCCAGTGGTGGTGGAGATTAATCTTCGCTGGCGGCTCTG 120
QY 1232 CACTGGAGAAGTCCGGTCTGCAACTGGTGGCGTCCGTAATCCCTATGAAGAGCGAAGA 1291
DB 121 CCTGGGAAGTCGAGATGTACAAATGGTGAATGATGTCTCTGCGTGGGAAGAGATGAAC 180
QY 1292 TTCGATATCTTAATCTTCACAGATGTCATCGCTGGCAGGTAGCTTAATCGGTCARA 1351
DB 181 TCGGATGCTTAATGGCAGCACTCTTTCTCGCTTATCTGGGTTACCTCTCAGGATCG 240
QY 1352 AATATATCACCAAGCACAATGACCGATTTTATCTATCAGATTTGCCGCCCTACGTGA 1411
DB 241 CCCATATCAGTATGTCATGCGAGGATCGCGCATTTTCGCCATGCGCGCAGCAATTAATGC 300
QY 1412 CGGAAGATGTCATCTTCTGTTGGCGATPACGGTATCGATTGGCAACCTACCGGTGATG 1471
DB 301 TGGATGAGCAAGCGCGACACTGCAAAATTAAGATGTCGATTTAACACAATATCGGATA 360
QY 1472 TTGTACTCAAGCGTTTACCAATCCACATATTCAGGACACCAACCGCTGCGTGGCG 1531
DB 361 AGTTAATTCACAGCTTTTGTGTAATCCGGCTGAAACATAAGACCTGGCAATCGCGATGG 420
QY 1532 ATGGTTTCTCGAAATTCGG 1551
DB 421 ATGGCAGCCAGAAATTAACCG 440

RESULT 4
CNS01QW5/c
LOCUS

DEFINITION
Anopheles gambiae GSS SP6 end of clone 31M20 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.

ACCESSION
VERSION
KEYWORDS

SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

CNS01QW5
Anopheles gambiae GSS SP6 end of clone 31M20 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.

ACCESSION
VERSION
KEYWORDS

SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

DEFINITION	28a07 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium sp. NGR234 genomic clone 28a07, DNA sequence.							
ACCESSION	AZ578838							
VERSION	A2578838.1	GI:11606521						
KEYWORDS	GSS.							
SOURCE	Rhizobium sp. NGR234.							
ORGANISM	Rhizobium sp. NGR234 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium. 1 (bases 1 to 244)							
REFERENCE	Viprey, V., Rosenthal, A., Broughton, W.J. and Perret, X. Genetic snapshots of the Rhizobium species NGR234 genome Genome Biol. 1 (6), RESEARCH0014 (2000) 21114532							
TITLE	Contact: Virginie Viprey							
JOURNAL	Laboratoire de Biologie Moleculaire des Plantes Superieures							
MEDLINE	University of Geneva							
COMMENT	1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland Tel: +44(0)1603450005 Fax: +44(0)1603450045 Email: virginie.viprey@bbsrc.ac.uk Class: shotgun.							
FEATURES	Location/Qualifiers							
source	1..244							
	/organism="Rhizobium sp. NGR234"							
	/strain="ANU265"							
	/db_xref="taxon:394"							
	/clone="28a07"							
	/clone_lib="Shot-gun genomic library of Rhizobium strain ANU265"							
	/notes="Vector: M13; derivative strain of NGR234 cured of pNGR234a"							
BASE COUNT	42 a	84 c	76 g	41 t	1 others			
ORIGIN								
Query Match	2.1%;	Score 39.2;	DB 17;	Length 244;				
Best Local Similarity	49.7%;	Pred. No. 0.69;						
Matches	98;	Conservative	0;	Mismatches 99;	Indels 0;	Gaps 0;		
QY	1045	GTGATCGACTGGCTGAGTACAATAATACCACATTGCCCGGAATACCATGGTTGACCGCATTACG	1104					
Db	48	GACCTGGCCGCCTGGCGCAAAAGTATACGTCGCCCTCCACCATGGTCGACCGGATCGTG	107					
QY	1105	CCTCGTCGGCAGCAGAACCTTCGGGCACGGATCAAGGCTCAAACGGGTATTGCCGATAAA	1164					
Db	108	CGGCCACACGGACACGCGACCGCGCGATCGACGCTTCGCTCGGCTCGAGGACGCC	167					
QY	1165	CGCCGGTATGGCGGAACCTTTATCCAGTGGGCTGTGGAAGATATTTCCGTGATGTC	1224					
Db	168	TGGCGCATCATGACCGAACCTTCCGGCAATGGGTGATCGAGGACGATTTCCCCTCGGT	227					
QY	1225	CGTCCGGCACTGGAGAA	1241					
Db	228	CGCCCCGCGCTGGCANGA	244					
RESULT 6								
AQ758091/c								
LOCUS	AQ758091		391 bp	DNA	linear	GSS 27-JUL-1999		
DEFINITION	H5_5552_A1_A07_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1128 Col=13 Row=A, DNA sequence.							
ACCESSION	AQ758091							
VERSION	AQ758091.1	GI:5623119						
KEYWORDS	GSS.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS	1 (bases 1 to 391)							
	Maharaj, G.C., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.							
TITLE	Sequence-tagged connectors: A sequence approach to mapping and							

scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 99380589

COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1128 row: A column: 13
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 391.
 Location/Qualifiers
 1..391
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1128 Col=13 Row=A"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"
 143 a 59 c 74 g 114 t 1 others
 BASE COUNT
 ORIGIN

Query Match 2.08; Score 37.8; DB 17; Length 391;
 Best Local Similarity 50.88; Pred. No. 2.4;
 Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1324 GCCTGGGCAGGTACGTTAAATCGGTCAAAATATATATCCACGAAAGCACAAATGACCGATTTT 1383
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 295 GCCTTGACTAGTACTTTTGATTTTACCTTAAATTTGTGATCTATTTACCTACAAATGTTCAAGTTT 236
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1384 ATCTATCAGATTTGCCGACCGCTAGTGCAGGAAAGATGTCATTCCTGTTGGCGGATAAC 1443
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 235 TTCCTTAAGATAGATTCCTGATGTGAGATGCTTGAAGTCATGGCATCAACATGCTAAGTT 176
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 1444 GGATATCGATTTGCCAAGCTACCGGTGATGTGTACTCAAGCGTTTACCACATCCACAT 1500
 1 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 175 TGGGACACATGCTCAAGCTCCTTTCCAGAAAGACTGTACCAATTTATCTATATACTT 119
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
 BG922047 632 bp mRNA linear EST 05-JUN-2001
 LOCUS BG922047.1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:495094 5',
 DEFINITION mRNA sequence.
 BG922047
 VERSION BG922047.1 GI:14302523
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 632)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rgs@bbs.femail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

10

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1M10906 row: e column: 17
 High quality sequence stop: 322.

FEATURES

Location/Qualifiers
 1..632
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="NCI_CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT
 ORIGIN

191 a 128 c 172 g 139 t 2, others

Query Match 2.0%; Score 37.6; DB 13; Length 632;
 Best Local Similarity 54.5%; Pred. No. 3.7;
 Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 490 CAATTACATGCTGCACATCGGTTAGTTCCTTTTCATCGGCACATCAGCGTGGTAT 549

DB 487 CACTTCGCATATTACTCTTGCTTAACCTTCATTGAACACGAAAGACTGTCGTAGTAT 546

QY 550 CTACACCGTTTCGAGGTGGCGGATAAAGCTGGAGCATTCGTCGGGGCAATATTCGT 609

DB 547 TGACACACTGCGTCGTCGAGCGCGGNCACCCGTGCGGAGCAGCGGGAACCTGGT 606

QY 610 AATGATGCTGAACA 623

DB 607 GAAAGGGCGCAC 620

RESULT 8
 A2115205/c

LOCUS A2115205 624 bp DNA linear GSS 12-MAY-2000
 DEFINITION RPCI-23-460F24.TV RPCI-23 Mus musculus genomic clone RPCI-23-460F24
 , DNA sequence.

ACCESSION A2115205
 VERSION A2115205.1 GI:7776066

KEYWORDS GSS.
 SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 624)
 Zhao, S., Nierman, M.G., Feldblyum, T., Malek, J., Shatsman, S., Akinret
 , B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)
 Other GSSs: RPCI-23-460F24.TJ

CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200

Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 460 row: F column: 24

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers
 1..624
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-460F24"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 148 a 161 c 129 g 186 t

ORIGIN

Query Match 2.0%; Score 37.2; DB 17; Length 624;
 Best Local Similarity 63.3%; Pred. No. 4.8;
 Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 967 TCGGATACGTCGCGCAATGTTGTAACGTTTCCATGATGCGCTGCTTTCAGTTCTCCAG 1026

DB 589 TCGGGATCTTCTCCAGATGTGTGAATTCACACAGGGCTGCCTTATCTTGTCCAG 530

QY 1027 CTAACTGGCAACAGGATGTCAATCGACTGG 1056

DB 529 CTCCCTGGCAGACGACAGTCATTCAGAGG 500

RESULT 9
 CNS010B7

LOCUS CNS010B7 787 bp DNA linear GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
 BACN03M07 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL098749
 VERSION AL098749.1 GI:5610360

KEYWORDS GSS.
 SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 787)
 Genoscope.

AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;

JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES

Location/Qualifiers
 1..787

/organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN03M07"

/clone_lib="DrosBAC"
 /plasmid="pBelOBAC11"

/note="end : T7"

BASE COUNT 52 a 114 c 73 g 19 t 529 others

ORIGIN

Query Match	2.08;	Score 36.4;	DB 12;	Length 550;
Best Local Similarity	55.6%;	Pred. No. 8;		
Matches	70;	Conservative 0;	Mismatches 56;	Indels 0; Gaps 0;
<hr/>				
Qy	981	CCATAATGGTCAACGTTTCCTCATGTAGCGCTGGTTGAGTTTCTCCAGCTFAACTGGCAAACA	1040	
Db	463	CCACAAAGTGCCCTTCTCCAGAAGGTGC'TGATGATGTTATTCACGTACGCCGGTAAGA	404	
Qy	1041	GGATGTCATCAGCTGGCTGAGTACAAAATPACCATTGCCCGAATACCATTGTTGACCCGAT	1100	
Db	403	GGATATCTACACCCGCTGATTTCTCTGAGCCCACTGGGAAGATAAGATGGATAAGAGCAT	344	
Qy	1101	TACGCC 1106		
Db	343	CCTGCC 338		
<hr/>				
RESULT 11				
AW165487/c				
LOCUS	AW165487	581 bp	mRNA	linear EST 10-NOV-1999
DEFINITION	G18047B01.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA sequence.			
ACCESSION	AW165487			
VERSION	AW165487.1	GI:6342663		
KEYWORDS	EST.			
SOURCE	Zea mays.			
ORGANISM	Zea mays			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.			
REFERENCE	1 (bases 1 to 581)			
AUTHORS	Walbot,V.			
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 618047 row: B column: 01. Location/Qualifiers 1. .581 /organism="Zea mays" /cultivar="Ohio43" /db_xref="taxon:4577" /clone_lib="618 - Inbred Tassel cDNA Library" /tissue_type="tassel" /dev_stage="tassel length from 0.1 to 2.5 cm" /lab_host="XL0LR" /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybridzap); Inbred tassel library from Schmidt lab"			
FEATURES	source			
	BASE COUNT	136 a	148 c	109 g 188 t
	ORIGIN			
<hr/>				
Query Match	2.08;	Score 36.4;	DB 10;	Length 581;
Best Local Similarity	55.6%;	Pred. No. 8.2;		
Matches	70;	Conservative 0;	Mismatches 56;	Indels 0; Gaps 0;
<hr/>				
Qy	981	CCATAATGGTCAACGTTTCCTCATGTAGCGCTGGTTGAGTTTCTCCAGCTFAACTGGCAAACA	1040	
Db	481	CCACAAAGTGCCCTTCTCCAGAAGGTGC'TGATGATGTTATTCACGTACGCCGGTAAGA	422	
Qy	1041	GGATGTCATCAGCTGGCTGAGTACAAAATPACCATTGCCCGAATACCATTGTTGACCCGAT	1100	
Db	421	GGATATCTACACCCGCTGATTTCTCTGAGCCCACTGGGAAGATAAGATGGATAAGAGCAT	362	
Qy	1101	TACGCC 1106		
Db	361	CCTGCC 356		

```

RESULT 12
AW171987
LOCUS      AW171987      619 bp      mRNA      linear      EST 15-NOV-1999
DEFINITION 618047B01.y1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
sequence.
ACCESSION  AW171987
VERSION     AW171987.1      GI:6431783
KEYWORDS    ze mays.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 619)
REFERENCE   Walbot,V.
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618047 row: B column: 01.
            Location/Qualifiers
                1..619
                /organism="Zea mays"
                /cultivar="Ohio43"
                /db_xref="taxon:4577"
                /clone_lib="618 - Inbred Tassel cDNA Library"
                /tissue_type="tassel"
                /dev_stage="tassel"
                /lab_host="xL0LR"
                /note="Organ: tassel; Vector: PAD-GAL4-2.1 (Hybrizap);
                Inbred tassel library from Schmidl lab"
BASE COUNT  217 a 106 c 189 g 105 t
ORIGIN
Query Match      2.0%; Score 36.4; DB 10; Length 619;
Best Local Similarity 55.6%; Pred. No. 8.5;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY  981 CCATAATGGTGAACGTTTCCATGATGGCTGTTGATGTTCTCCAGCTAACTGGCAACA 1040
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   309 CCACAAGGTGCTTCTTCCAGGAGGTCTGATGATGTTATCCAGTCAGCGGTAAGA 368

QY  1041 GGATGTCATCGACTGCTGCTGAGTACAAATACCACTTGCCTCCGCAATACATGTTGACCGCAT 1100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   369 GGATATCTACACCGCTGATTTCTCTGAGCCCACTGGAGAAGATAAGATGATAAGAGCAT 428

QY  1101 TACGCC 1106
      |||
Db   429 CTGCC 434

RESULT 13
AQ500296
LOCUS      AQ500296      771 bp      DNA      linear      GSS 29-APR-1999
DEFINITION V4264 mtN-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', DNA sequence.
ACCESSION  AQ500296
VERSION     AQ500296
KEYWORDS    GSS.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
            1 (bases 1 to 771)
REFERENCE   Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symonlati,D., Jansen,R.,

```

```

Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mtN-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTCTTGTGGAAGTAC
Class: transposon-tagged.
FEATURES
    source
        1..771
        /organism="Saccharomyces cerevisiae"
        /db_xref="taxon:4932"
        /clone_lib="mtN-3xHA/lacZ Insertion Library"
        /lab_host="E. coli"
        /note="Vector: pHS56-Sal; A yeast genomic DNA library
        (lacking mitochondrial DNA) was prepared in pHS56-Sal;
        genomic DNA was size-fractionated (DNA of roughly 2-3 kb
        in length) prior to cloning. This library was
        subsequently mutagenized with a mtN-3xHA/lacZ
        minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT  175 a 205 c 141 g 250 t
ORIGIN
Query Match      2.0%; Score 36.4; DB 17; Length 771;
Best Local Similarity 53.5%; Pred. No. 9.6;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY  1506 GCACACCAACCAACGCTGCTCGGATGGTTTCTGAAATTCGGCGCATGATTCGCC 1565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   280 GGAATAATATCAGCGCGCGCGCGCGGGTGGCGAACAACATCCAGCATGAGATTCCC 339

QY  1566 CACACTGCGGAGATGCTACACGAGCGGCTTCCGCCGAATGCCACGCCCATGTTACTGTC 1625
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   340 CGCGTGGGATGATTTCTCCAGCGCGGTGTTCCCGGAAACAGATTCCGGAGATCCTCA 399

QY  1626 ACTGTTTACGATTCATCGAG 1647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   400 ACTCTTTCCATTATAAAGGCG 421

RESULT 14
AY110366/c
LOCUS      AY110366      791 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays CL67890_1 mRNA sequence.
ACCESSION  AY110366
VERSION     AY110366.1      GI:21214775
KEYWORDS    HTC.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 791)
REFERENCE   Hayney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
            2 (bases 1 to 791)
JOURNAL     Coe,E.C.
REFERENCE   Direct Submission
AUTHORS     Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
JOURNAL     Location/Qualifiers
            1..791
            /organism="Zea mays"

```

```
/db_xref="MaizeDB:632757"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT      154 a   193 c   129 g   230 t      85 others
ORIGIN

Query Match      2.0%; Score 36.4; DB 11; Length 791;
Best Local Similarity 55.6%; Pred. No. 9.8;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 981 CCATAATGGTGAACGTTCCATGATGCGCTGGTTGAGTTTCTCCAGCTAACTGGCAAACA 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 CCACAAAGGTGCTTCTCCAGGAAGTGCTGATGATGTTATCCAGTCAGCGGGTAAAGA 426

QY 1041 GGATGTCATCGCTGCTGAGTACATACCACTTCCCGCAATACCATGTTGACCGCAT 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 GGATATCTACACCGGTGATTCTCTGAGCCCACTGGAGAAGATAAGATGGATAAGAGCAT 366

QY 1101 TAGGCC 1106
    |||
Db 365 CCTGCC 360

RESULT 15
AZ163357
LOCUS
DEFINITION
SP_0073_B2_A10_T7A Strongylocentrotus purpuratus, purple sea urchin
, sperm genomic BAC library Strongylocentrotus purpuratus genomic
clone Plate=73 Col=20 Row=B, DNA sequence.
ACCESSION
VERSION AZ163357.1 GI:8316057
KEYWORDS
SOURCE
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 533)
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 73 row: B column: 20
Seq primer: T7
Class: BAC ends
High quality sequence stop: 533.
Location/Qualifiers
1..533
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=73 Col=20 Row=B"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
```

```
BASE COUNT      192 a   95 c   103 g   135 t      8 others
ORIGIN
DH10B"

Query Match      1.9%; Score 36; DB 17; Length 533;
Best Local Similarity 48.1%; Pred. No. 10;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 96 TAAACGCTGGAGCATTGCTGCGGGCAATATTCGTAAATGATGCTGAACATCTCGTACAGGC 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 TAAGCAGGCTGGATCAATACGTGCCAATTCATCAAAATATTTGCATATTTGNTAATAAA 343

QY 156 ACTCAGTGCACAGAAAGGTGCTATGTGCTGGAACCGTCAGCCCGGAAGGGTAAAGCGA 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 TAGAGCATCACAGAAATGTCACCAAGGATGGGAAGAGAGAAGAGAAATGCTCAAGAGA 403

QY 216 ATATGAAGAGATCACCTCAATTCAGAGTTGATACCGTGGCAGGCAGATTTACAACGGCT 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 AGATGGAGAGATGACCTTAACATCATACATGGGAACCGCCGCTTGGACTANGATAGCTAGA 463

QY 276 GATTGCTGGAAGGGCGAGATCCGAAGA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 GACAGACGAGAAATGGAAATGTTATGA 489

Search completed: March 21, 2003, 23:49:07
Job time : 2848.31 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:06:57 ; Search time 11422.3 Seconds
(without alignments)
10117.650 Million cell updates/sec

Title: US-09-802-208B-2
Perfect score: 3971
Sequence: 1 atcattgagcagttgtt.....tcacgcgcgcagcatcgat 3971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3971	100.0	3971	1	AY005817 Escherich
2	3971	100.0	3971	6	AX268027 Sequence
3	1530.6	38.5	3233	1	AF045244 Klebsiell
C 4	824.8	20.8	10789	1	AE009731 Brucella
C 5	780.2	19.6	5930	1	AF045245 Klebsiell
C 6	643.2	16.2	10295	1	AE008252 Agrobacte
C 7	643.2	16.2	11437	1	AE009360 Agrobacte
C 8	468.6	11.8	851	1	KPNRBTDK M25606 K.aerogenes
C 9	369	9.3	10201	1	AE013623 Versinia
10	369	9.3	235050	1	AJ141458 Versinia
11	302.4	7.6	1923	9	BC014947 Homo sapi
12	300.8	7.6	1991	6	AX427495 Sequence
13	282.6	7.1	2485	9	AK090568 Homo sapi
14	226	5.7	1842	10	BC031708 BC031708 Mus muscu
15	218.6	5.5	11164	1	AE009374 Agrobacte
C 16	218.6	5.5	14945	1	AE008240 Agrobacte
C 17	205.8	5.2	2858	3	AY060286 Drosophil
C 18	204.8	5.2	323450	1	SME591790 AL591790 Sinorhizo
C 19	183	4.6	1947	9	AK022237 Y15896 Bacillus su
C 20	179.2	4.5	17787	1	BSY15896 AC015395
C 21	177.4	4.5	45258	2	AC015395 AC015395
C 22	177.4	4.5	87500	1	AF027868 Bacillus
C 23	177.4	4.5	186218	3	AC023680 Drosophil
C 24	177.4	4.5	310493	3	AE003480 Drosophil
C 25	172.2	4.3	1642	9	AK054842 Homo sapi
C 26	169.6	4.3	2664	1	BSBOFCEN X93081 B.subtilis
C 27	134.4	3.4	1361	1	KARBT X02448 Klebsiella
C 28	124	3.1	1044	9	AK001848 Homo sapi
C 29	120.2	3.0	1180	9	BC000610 Homo sapi
C 30	104	2.6	38813	8	SC9727 Z48758 S.cerevisia
C 31	93.2	2.3	20456	1	AE008908 AF086781 Rhizobium
C 32	91.4	2.3	184	1	AF086781 Rhizobium
C 33	91	2.3	329709	1	AP002997 Mesorhizo
C 34	90.4	2.3	677	6	AX385150 Sequence
C 35	75.8	1.9	39446	1	SC87 ALI32674 Streptomy
C 36	75.4	1.9	696	8	CNS010X8 ALI16100 Botrytis
C 37	73.8	1.9	1124	8	BNA243083 AJ243083 Brassica
C 38	72.8	1.8	10127	1	AE010077 Streptoco
C 39	72.6	1.8	10592	1	AE013686 Versinia
C 40	72.6	1.8	22050	1	AJ141456 Versinia
C 41	72.4	1.8	10346	1	AE005436 Escherich
C 42	72.4	1.8	268857	1	AP002560 Escherich
C 43	72.2	1.8	297850	1	AP003137 Staphyloc
C 44	72.2	1.8	303249	1	AP001515 Bacillus
C 45	72.2	1.8	341350	1	AP003365 Staphyloc

ALIGNMENTS

RESULT 1
AY005817
LOCUS
DEFINITION
Escherichia coli ribitol dehydrogenase (rtld), ribitol kinase (rtlk), and ribitol transporter (rbtr) genes, complete cds.
ACCESSION
AY005817
VERSION
AY005817.1
KEYWORDS
GI:11036624
ORGANISM
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 3971)
AUTHORS
Lafayette,P.R. and Parrott,W.A.
TITLE
A non-antibiotic marker for amplification of plant transformation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
vectors in E. coli
Plant Cell Rep. 20, 338-342 (2001)
2 (bases 1 to 3971)
LaFayette,P.R. and Parrott,W.A.
Direct Submission
Submitted (26-Jul-2000) Crop and Soil Sciences, University of
Georgia, Plant Sciences Bldg., Athens, GA 30602, USA
Location/Qualifiers
1..3971
/organism="Escherichia coli"
/strain="C"
/db_xref="taxon:562"
96..848
/gene="rtld"
96..848
/gene="rtld"
/codon_start=1
/transl_table=11
/product="ribitol dehydrogenase"
/protein_id="AA01883.1"
/db_xref="GI:11036625"
/translation="MMHNSVSPMNTPLNGKVAATGAASGIGLOCAKTLDDAGAKVVL
IDREGDLKHIVAEIGENAYALQDLFNQOVDNMLADIIELAGSLDIFHANAGAYIG
CPVARGDPQWDRVLNINARFCYRAVLPHMIAQRSGDIIFTSSIAGVVPVINEPI
YIASFAVQAFVHTTRQVSQYGVRAVLPGVPVIALDDMPKAMEEALANGSLMQ
PIEVAESVLFMWTRSKNVTVDLVILPGSVDL"
859..2463
/gene="rtlk"
859..2463
/gene="rtlk"
/codon_start=1
/transl_table=11
/product="ribitol kinase"
/protein_id="AA01884.1"
/db_xref="GI:11036626"
/translation="MTIKTVIGVDVSGSVRAGIFDLNLSLLSHATEKITTRSGS
RVEQSSEIHWQVCIRNALTDVCAQSVAGIFDQATSLVLDKNGDPLVPSPFG
DAKQNIIVMDHRTAEQERINATHHPVLNIVYGGKISPEMETPKILMLKENMPEIYER
AQGFIDLHATRWATGLARSCTVCTCKTWLAHENRMDPDYFTIIGLAELADEDFI
RIGHIVSPGTCGNGLAQAAEMGLPGTPVAVGLIDAHAGGIGTVGVEGALNNL
AVVFTSGTMASTSPSFVPGVWPYTSAMVPLGLVVEGGSAAAGAIIDOLLDPHA
VEAREAKORNPPLVWLADRIEKTQPSDAVALAKGLHVVPFELGNRAPFADPHA
RAVIGLGMERDLNLLALYIAGLCGIGYGLRQILDQAQTAQVSNKVINISVGAGQHP
LVKQILADTGCPVITTOCEPVLIGLSAILGAVAGNIAPSVGEAMQQTTHVDKYITYP
ERYQLSHRRYRAYQLQHTAKLRD"
2565..3839
/gene="rbt"
2565..3839
/gene="rbt"
/codon_start=1
/transl_table=11
/product="ribitol transporter"
/protein_id="AA01885.1"
/db_xref="GI:11036627"
/translation="MSRNNKQWLGPLHLIWCYIAIAVMTGDGFELAFLSHYIKSLG
FTPACASFTLYGLAALSAWSGVVAEIIIPQKMLIGFVLMCVFVLFVFGGQ
ANYGILIFGIRGLAYPLFLYFTVIHNRVSRSSSALGWYAVYVSGVAGSY
IPSFPIPMGENTLWALAFCFAGGVIAIISLRHVKTGPHMHNLTPREKFAELSRV
TLLYNIRNIFLSSIVRIINTLSLFGFVIMPMWFDELGFTTSEWLOYWAAFFFTTIF
SNIFWGIYAEKGMWRVIRWFGLGMASSLAFLYMPQYFGHNYWMIIPALALGTFV
AASFVMAVFPALFKHKAISVNLNLSAGMSNLAPALAVVLLPWFSTIGVVIAYTA
LYLLAFVLCAFIRVQPGFSSAPVTEKALNIS"
BASE COUNT 919 a 930 c 1020 g 1102 t
ORIGIN
Query Match 100.0%; Score 3971; DB 1; Length 3971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCGATTGACGCTTTCACACGGCAGCTAAATCCCGTTTCAGTCGCGTCAACGCAAA 60
Db 1 ATCGATTGACGCTTTCACACGGCAGCTAAATCCCGTTTCAGTCGCGTCAACGCAAA 60

QY 61 CAGAGACTATAAATTCGCCCTCGTAAAGGATTATATGATGAATCACTCTGTGCCCTCTA 120
Db 61 CAGAGACTATAAATTCGCCCTCGTAAAGGATTATATGATGAATCACTCTGTGCCCTCTA 120
QY 121 TGAATACTCCCTTAAATGCAAAAGTTGCAGCTATCACTGGCGCTGCGTCAGGTATGGCC 180
Db 121 TGAATACTCCCTTAAATGCAAAAGTTGCAGCTATCACTGGCGCTGCGTCAGGTATGGCC 180
QY 181 TGAATGTGCAAAACGCTGCTCGATGCAGGACAAAGTAGTATTGATTGACCCGGAAG 240
Db 181 TGAATGTGCAAAACGCTGCTCGATGCAGGACAAAGTAGTATTGATTGACCCGGAAG 240
QY 241 GCGACAACTGCACAAAGATTGCTGCTAGTATAGCGGAAACGCGTACGCGCTGCAACTCG 300
Db 241 GCGACAACTGCACAAAGATTGCTGCTAGTATAGCGGAAACGCGTACGCGCTGCAACTCG 300
QY 301 ATCTCTTCAATATCAGCAAGTCGATAAATCTGCGGACATATTCGAACCTGGCGGGTG 360
Db 301 ATCTCTTCAATATCAGCAAGTCGATAAATCTGCGGACATATTCGAACCTGGCGGGTG 360
QY 361 GGCTGGATATTTTTCATGCCAATGCAGGCGCTTATATTGGCGGCCAGTGGCTGAAGGTG 420
Db 361 GGCTGGATATTTTTCATGCCAATGCAGGCGCTTATATTGGCGGCCAGTGGCTGAAGGTG 420
QY 421 ATCCAGATCTGGGATCGTGTGTTAAATCTGAATATAAATGGGCGCTTCGCTGTGTC 480
Db 421 ATCCAGATCTGGGATCGTGTGTTAAATCTGAATATAAATGGGCGCTTCGCTGTGTC 480
QY 481 GTGCGAGTCTGGCGCATATGATTGCGCAGAGTCGCGGCATATAATTTTACCAGTTCCA 540
Db 481 GTGCGAGTCTGGCGCATATGATTGCGCAGAGTCGCGGCATATAATTTTACCAGTTCCA 540
QY 541 TCGCGGCGCTCGTTCCGGTTATCTGGGAAACCGATCTACACCGGTCCTCAAAATTTGCCGTTT 600
Db 541 TCGCGGCGCTCGTTCCGGTTATCTGGGAAACCGATCTACACCGGTCCTCAAAATTTGCCGTTT 600
QY 601 AGCATTCGTACACACTACCGCGCCAGGTTTCTCAATATGCGCGTGGTGGGTGGG 660
Db 601 AGCATTCGTACACACTACCGCGCCAGGTTTCTCAATATGCGCGTGGTGGGTGGG 660
QY 661 TGTGTCAGGACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 661 TGTGTCAGGACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 721 AAGCCCTGGCAAAATGATGCTGATCAACCGATTCGAGTGGCGGAATCAGTATTGTTA 780
Db 721 AAGCCCTGGCAAAATGATGCTGATCAACCGATTCGAGTGGCGGAATCAGTATTGTTA 780
QY 781 TGGTGACCGCTCGAAAAATGTCACCGTGGGAGATTTAGTGATCCTGCTGGCAGTCTCG 840
Db 781 TGGTGACCGCTCGAAAAATGTCACCGTGGGAGATTTAGTGATCCTGCTGGCAGTCTCG 840
QY 841 ATCTGTAAGGGCGCAATCATGACAATAGCAAAACCGTTATTGGTGTGATGGGATCA 900
Db 841 ATCTGTAAGGGCGCAATCATGACAATAGCAAAACCGTTATTGGTGTGATGGGATCA 900
QY 901 GGCAGTGTCCGCGCCGGGATTTTGTATCTCAACCGATCTCTGCTATCCCATGCCAGAA 960
Db 901 GGCAGTGTCCGCGCCGGGATTTTGTATCTCAACCGATCTCTGCTATCCCATGCCAGAA 960
QY 961 AAATCAGACTACGCGCGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCTGG 1020
Db 961 AAATCAGACTACGCGCGCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCTGG 1020
QY 1021 CAGCGGCTGTGTCATGATTCGTAATGCGCTCACTCTGCGACAGCTTTGTGCACAAAGT 1080
Db 1021 CAGCGGCTGTGTCATGATTCGTAATGCGCTCACTCTGCGACAGCTTTGTGCACAAAGT 1080
QY 1081 GTGGCAGGATCGGTTTGTGATGCCACCTGTTCTCTGGTGGTACTGGATAAAACGGTGTAT 1140
Db 1081 GTGGCAGGATCGGTTTGTGATGCCACCTGTTCTCTGGTGGTACTGGATAAAACGGTGTAT 1140
QY 1141 CCATTGCGCTGTGAGCGCCGGAAGAGATGCAAAAGAAATATCATTTGTGTGGATGCATCAC 1200

1021 CAGCGGCTCTGTTTCATGTTATTCGTAATCGCTCACTCTGGCAGAGCTTTGTGCACAAAGT 1080
1021 CAGCGGCTCTGTTTCATGTTATTCGTAATCGCTCACTCTGGCAGAGCTTTGTGCACAAAGT 1080
1081 GTGGCAGCATCGGTTTGTATGCGCACCTGTTCTCTGTGGTGACTGAGATAAAGCGGTGAT 1140
1081 GTGGCAGCATCGGTTTGTATGCGCACCTGTTCTCTGTGGTGACTGAGATAAAGCGGTGAT 1140
1141 CCATTGCTCTGTCAGCCCGGAAGGAGATGCAAGCAAAATATCATTTGCTGATGGATGATC 1200
1141 CCATTGCTCTGTCAGCCCGGAAGGAGATGCAAGCAAAATATCATTTGCTGATGGATGATC 1200
1201 CGGCCACCGAAGCAGAGCATCAATGCCACTCACCATCCGGTCTGAACTACGTC 1260
1201 CGGCCACCGAAGCAGAGCATCAATGCCACTCACCATCCGGTCTGAACTACGTC 1260
1261 GGTGGTAAATTTCCGCTGAAATGGAACACCGAAATTTCTGGCTGAAAGAAATATG 1320
1261 GGTGGTAAATTTCCGCTGAAATGGAACACCGAAATTTCTGGCTGAAAGAAATATG 1320
1321 CCAGAGATCTACGACGTGCGGACAATTTTTCGATCTGGCGGATTTCTGACCTGGCGG 1380
1321 CCAGAGATCTACGACGTGCGGACAATTTTTCGATCTGGCGGATTTCTGACCTGGCGG 1380
1381 GCTACCGGTGATTTAGCCGCTTCAGTATGCATGTTACCTGTAAATGGACGTGGCTGGCA 1440
1381 GCTACCGGTGATTTAGCCGCTTCAGTATGCATGTTACCTGTAAATGGACGTGGCTGGCA 1440
1441 CATGAAATCCGCTGGATCCAGATTTATTCGCCACCATTTGSCCTTGACAGTTAGCGGAT 1500
1441 CATGAAATCCGCTGGATCCAGATTTATTCGCCACCATTTGSCCTTGACAGTTAGCGGAT 1500
1501 GAAATTTTATTCGATTTGGTTCATCATATTTCTCCGGAACACCTTTGTGGAATGGT 1560
1501 GAAATTTTATTCGATTTGGTTCATCATATTTCTCCGGAACACCTTTGTGGAATGGT 1560
1561 TTAACAGACAAAGCGCGGACAGATGGATTTACTCCCGGACACACCTGCTGCTAGGG 1620
1561 TTAACAGACAAAGCGCGGACAGATGGATTTACTCCCGGACACACCTGCTGCTAGGG 1620
1621 TTAATGATGCTACGCTGGTGGCATGTTGCTGCGGTAGAGGTGGAGCGCTGAAC 1680
1621 TTAATGATGCTACGCTGGTGGCATGTTGCTGCGGTAGAGGTGGAGCGCTGAAC 1680
1681 AATCTCGCGTATGTTTCGCGCATCTCTCATGCACCATGGCATCTACCACTTCTCCCTCG 1740
1681 AATCTCGCGTATGTTTCGCGCATCTCTCATGCACCATGGCATCTACCACTTCTCCCTCG 1740
1741 TTTGTACCGGGTGTCTGGGGCCGATTTACAGTGGATGGTTCCAGGTCTGFGTTAGTT 1800
1741 TTTGTACCGGGTGTCTGGGGCCGATTTACAGTGGATGGTTCCAGGTCTGFGTTAGTT 1800
1801 GAAAGCGGCAAGGTCTGCGGAGCAGCTATTGACAGCTACTTGTATTCATCCCGCT 1860
1801 GAAAGCGGCAAGGTCTGCGGAGCAGCTATTGACAGCTACTTGTATTCATCCCGCT 1860
1861 GTTGAAGAGCTCGGAAATGGCAACAGTGTGAATCAGCCCTCCCGCTCTGGCTTGGT 1920
1861 GTTGAAGAGCTCGGAAATGGCAACAGTGTGAATCAGCCCTCCCGCTCTGGCTTGGT 1920
1921 GATCGAATCTCGAANAACGGCGCAACCATCAGATGCTGCGCCCTGGCGAAGGGGTA 1980
1921 GATCGAATCTCGAANAACGGCGCAACCATCAGATGCTGCGCCCTGGCGAAGGGGTA 1980
1981 CACGTGGTGGGATTTCTCGGAATCGGCCCTTCGAGATCCTCATGCCAGAGCG 2040
1981 CACGTGGTGGGATTTCTCGGAATCGGCCCTTCGAGATCCTCATGCCAGAGCG 2040
2041 GTAATTTCTGGCTGGTATGGAGCGAGATCTGGATAATTTACTCGCCCTTGTATATCGCT 2100
2041 GTAATTTCTGGCTGGTATGGAGCGAGATCTGGATAATTTACTCGCCCTTGTATATCGCT 2100

2101 GGATTATCGGAATGGTTATGGTCTGCCCAAAATTCGACGCTCAACACAGCGCAGGA 2160
2101 GGATTATCGGAATGGTTATGGTCTGCCCAAAATTCGACGCTCAACACAGCGCAGGA 2160
2161 GTAGTGAGTAAATAATCGTTATTTAGCGCGGTGCCGGCAGCATCCACTGTFAGACAA 2220
2161 GTAGTGAGTAAATAATCGTTATTTAGCGCGGTGCCGGCAGCATCCACTGTFAGACAA 2220
2221 ATTCTGGCAGATACCTCGCGGTATTCGGGTCAATTACACGCAATGTCGCAACCGGTTTA 2280
2221 ATTCTGGCAGATACCTCGCGGTATTCGGGTCAATTACACGCAATGTCGCAACCGGTTTA 2280
2281 TTAGCTCGGCCATTCCTTTGGTCTGTCGCCGGAATATTTGACCTTCTGTTGGCGAAGCG 2340
2281 TTAGCTCGGCCATTCCTTTGGTCTGTCGCCGGAATATTTGACCTTCTGTTGGCGAAGCG 2340
2341 ATGCACAATTTACCCCATGTCGATATAATTTATTTATTCGCAAGACGCTATCAATCTCTT 2400
2341 ATGCACAATTTACCCCATGTCGATATAATTTATTTATTCGCAAGACGCTATCAATCTCTT 2400
2401 CATCATCTCGATATGAGCGCTTATAGCAGTTGACAGCATCTGCAAAATTTACTCAGAGAC 2460
2401 CATCATCTCGATATGAGCGCTTATAGCAGTTGACAGCATCTGCAAAATTTACTCAGAGAC 2460
2461 TAAATTAACAGCGCGCTGACGCTGTTTTCAGGCAATCACTAATAACGACTACTCCGGTA 2520
2461 TAAATTAACAGCGCGCTGACGCTGTTTTCAGGCAATCACTAATAACGACTACTCCGGTA 2520
2521 ATATCCCGAGTGCAATTCATCTGACCCCTTAAAGAGAGTTTATATGTCGCAAAATAATA 2580
2521 ATATCCCGAGTGCAATTCATCTGACCCCTTAAAGAGAGTTTATATGTCGCAAAATAATA 2580
2581 AACAGTGGTTGGTTGCGCATCTGATATGGGATATATCGCCATCGCAGTATTTA 2640
2581 AACAGTGGTTGGTTGCGCATCTGATATGGGATATATCGCCATCGCAGTATTTA 2640
2641 TGACTGGTGGTTTTCGAACTCGCATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700
2641 TGACTGGTGGTTTTCGAACTCGCATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700
2701 CACCTGCGGAAGCTCTTTTGGCTTTACGCTCTAGCGCTGGCGCTGCCCTTTCCGCT 2760
2701 CACCTGCGGAAGCTCTTTTGGCTTTACGCTCTAGCGCTGGCGCTGCCCTTTCCGCT 2760
2761 GGGTTTCTGGGTAGTAGCGAAATCATCACGCCCAAAAAACCATGCTGATTGGTTTG 2820
2761 GGGTTTCTGGGTAGTAGCGAAATCATCACGCCCAAAAAACCATGCTGATTGGTTTG 2820
2821 TCCTATGGTGGTATTCATGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 2880
2821 TCCTATGGTGGTATTCATGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 2880
2881 GGTAAATCTGCTGTTTACGGGATTCGTTGTTTACGCTATCCGCTATTTCTCTACTCAT 2940
2881 GGTAAATCTGCTGTTTACGGGATTCGTTGTTTACGCTATCCGCTATTTCTCTACTCAT 2940
2941 TTATCTGCTGTTATTTATTAACAGTGGTAGCGAAATTTCCAGTTCTGCTGTTGGTGT 3000
2941 TTATCTGCTGTTATTTATTAACAGTGGTAGCGAAATTTCCAGTTCTGCTGTTGGTGT 3000
3001 ACTGGCGGTATATTCAGTAGGGATCGGTTGCTGGCAGTTATATTTCCAGTTTACGA 3060
3001 ACTGGCGGTATATTCAGTAGGGATCGGTTGCTGGCAGTTATATTTCCAGTTTACGA 3060
3061 TACCGATTATGGTCAAAATGGGAACCTTATGCTGGCAGTTCTGCTTTTGTGGGG 3120
3061 TACCGATTATGGTCAAAATGGGAACCTTATGCTGGCAGTTCTGCTTTTGTGGGG 3120
3121 GTGCTATTGCCATGATCTCTTGGCTCATGTTTAAACGCGCTGACATATGCAATAATTA 3180
3121 GTGCTATTGCCATGATCTCTTGGCTCATGTTTAAACGCGCTGACATATGCAATAATTA 3180
3181 CTCCCGGTGAGAAGTTTGCAGAAATTAAGTCGGGCGAGTAACCTTTATATACCAACCGCA 3240

Db 61 CCGTGGCGACATTGTGATCCTGCCGAACACGCGTGGATCTCTGACCCCTGGCCAGCCCTGTG 120
Qy 846 TAAGGGCCCAATCATGACAATAACCAAAACCGTTATTGGTGTGATGTGGGATCAGGCAG 905
Db 121 AGAGGATGTGATACGATACGATACGCAAAACATTTATTGGCGTGGATGTCGGTTCGGGCG 180
Qy 906 TGTCCGCGCGGGATTTTGTGATCTCAAGGGATCTCTGCTATCCCATGSCACAGAAAAAT 965
Db 181 CGTGGCGCGGGGTATTCAATCTCGGGGGCGAGCTCTCCGCCACGCGACCCGGGAAT 240
Qy 966 CACGACTACGGCGCGCAGCGAAAGCCGCTGGAACAGTCCAGCCAGGAGATCTGGCAGGC 1025
Db 241 AACCCGTGTTTCGACGCGCGGAAACAAAGTCGAGACGTCGAGCGGTGAGATCTGGCAGGC 300
Qy 1026 GGTCTGTTCATGTATTCGTAATGGCTCACCTCTGCGACAGCTTGTGACAAAGTGTGGC 1085
Db 301 GGTGTGCTACTGATTAACAAAGCGGGTGGCCAAAGCGCGCTTTCGCGCTGTCGATCGC 360
Qy 1086 AGGATCGGGTTTGTATGCCACTGTTCTCTGTGGTACTGGATRAAAACGGTGATPCAT 1145
Db 361 CGGATCGGGTTTCGATGCCACTGCTCGCTGGTGTGATGGCGATAACGATGGCGCGCT 420
Qy 1146 GCCTGTACGCCCGGAAGAGATGCAAGCAAAATATCATTTCTGTGGATGGATCAACCGGC 1205
Db 421 GCGGTGCGGCTTCGGGACGATGCGGACCGCAACATTTATGCTGGATGGATCAFCGCGC 480
Qy 1206 CACGAAACAGCAGAGCGGAATCAATGCCACTCACCATCCGGTGGTGAATACGTCCGTGG 1265
Db 481 CACGGGCGAGCGGAAAGATCAATGCCACGGGCCACCCGCTACTGCGGTACGTTGGGG 540
Qy 1266 TAAATTTCCGCTGAAATGGAAACACCGAAATTTCTCTGGCTGAAAGAAATATGCCAGA 1325
Db 541 CAAATCTCGCGGAGATGCAAAAGCGCGAAATTTGTGGCTGAAGGAGAAATCGCCGCCA 600
Qy 1326 GATCTACGAACGTCGCGGACAAATTTTCGATCTGCGCATTTCTGACCTGGCGGGCTAC 1385
Db 601 CATCTATCAGTGGCGGCCACTCTCTCGACCTCGCGGATTTACCTACCTGGCGCTCAAC 660
Qy 1386 CGGTGATTAGCGGTTTCAGTATGCACCTGTGTACCTGTAAATGGAGCTGGCTGGCACATGA 1445
Db 661 CGCGATGAAGCCGCTCGGTCTGCAGGTCGACCTGTAAATGGACCTACCTGGCGCATGA 720
Qy 1446 AAATCGCTGGATCAGATATTTCGCGACCATTTGGCCCTTCGACGATTAGCGGATGAAGA 1505
Db 721 ACAAGCTGGGATCGGGTTACTTCCGCGAGATTGGCTCGAAAGAGCTGGTGGATGAAGA 780
Qy 1506 TTTTATTCGCATGTGTCATCATATTGTTTCTCCGGAACCTTGTGGAATGTTTTAAC 1565
Db 781 TTTTGTCCGCATCGGCCAGCGGATTGTCGATCCAGGCACCCCTCGCGCGAGGGCCCTTG 840
Qy 1566 AGCAAGCCGCGCAGAGATGGGATTACTCCCGGCAACACTGTCTGCTGTAGGGTTAAT 1625
Db 841 CGCCAGCGCGCAGAGGATGGGCTTCGCGATCGGTACGCGGCTGGCGGTGGGATCAT 900
Qy 1626 TGATGCTCACGCTGGTGGCATCGGTACGGTTGGCGTAGAAGTGGAGCGCTGAACAATCT 1685
Db 901 TGAGCCCAACGCGCGGCATCGGTACCGTGGCGTGTAAACGCGCGGTCAACAATAT 960
Qy 1686 CGCGTATGTTTCGGCACTTCTCATGCACCATGGCATCTACCACTTCTCCCTGTTGT 1745
Db 961 GGCCTATGTCTTCGGCACTTCTCTCGACCATGACCAACCACTCAGGAAGCGGTATTCTG 1020
Qy 1746 ACCGGTGTCTGGGGCGGATTAACAGTGGCATGTTCCAGGCTCTGCTGTAGTTGAAGG 1805
Db 1021 CCGGGCGTCTGGGCCCTTACTATTCGGCGATGGTCCGGGCTACTGGCTAAGTGAAGG 1080
Qy 1806 CGGCAAAAGTCTCGGGGAGCAGTATTGACCACTACTTGATTTCCATCCGGCTGTTGA 1865
Db 1081 CGGCAAAAGCCCGCGCGCGCTATCGACCACTGCTGAGCTTCCATCCGCGCGCGC 1140
Qy 1866 AGAAGCTCGGAAATGGCAACAGTGAATACGCCCTCCCGTCTGGCTGTGCTGATCG 1925
Db 1141 GGAAGCGCGGAGCAGCGAAAGCGGAGGCTCCCTCTACCGGCTCTGCGTGGCGGATCG 1200

Qy 1926 AATCCTCGAAAAAAGCGCGCAACCATCAGATGCTGTGCGCCCTGGCGAAAGGCTACAGT 1985
Db 1201 CGTCTGACGCAAGTTGTTCCCTTCTGAGCGGTACGCTGGCGCGGCTCTGCAGT 1260
Qy 1986 GGTGCGGGAATTTCTCGGAAATCGGCCCTTCCAGATCTCTATGCGAGAGCGGTAAAT 2045
Db 1261 CGTTCGGAGTTTCTCGGTAATCGCGCGCGCTTCCGATCCCAACGGAAGCGCTGAT 1320
Qy 2046 TTGTGGCTGGGTATGGAGCGAGATCTGATAAATTTACTCGCTTGTATATCGCTGGATT 2105
Db 1321 TGCCGGGTGGGATGGAGCGGACCTGGATAAATCTGACCGGCTGTAGTTGCCGGGT 1380
Qy 2106 ATGGGGAATTTGTTATGCTCGCGCAAAATTTCTCACCTCAAAACAGGCGAGGAGTAGT 2165
Db 1381 ATGGGGAATTTGCTATGCTTGGACAGATCATTTGACGCCACGCGCTGCGGAATGA 1440
Qy 2166 GAGTAAATATCTGTTATTTAGGGGGGTGGCGGAGATCCACTGGTACGACAATTTCT 2225
Db 1441 GAGTAAATATCGTTCATCAGCGCGGAGCGGGGAGCATCGCTGGTGGCTGAGTGT 1500
Qy 2226 GCGAGATACCTCGGTTATTTCCGGTCAATACACGCAATGCTGCGAAACCGGTTTTATTAG 2285
Db 1501 GCGCGACGCTTGGCGGTAGGCTGCTGAGCACCCGCGAGCGCGGCTGCTGTAGG 1560
Qy 2286 CTCGGCCATTTCTGCTGCTGCGGAAATATTGCACTTTCTGTTGGCGAAGCATGCA 2345
Db 1561 TTCCGCCATCTCGCGCGGTGGCGGAGAGTCCGCGCTTCCGTCGCCGAGGCCATGA 1620
Qy 2346 ACAATTCACCATGTGGGATAAATATTATTTCCGCAAGACGCTATCAATCTCTTCATCA 2405
Db 1621 GCAGTTTACTCAGTGGATGCCACCTATCACAGCGAGCGGCTTTAGCCCGCTTCATCA 1680
Qy 2406 TCGTCGATATGAGGCTTATAAGCAGTTCGACGATCTCAAAATTTACTCAGAGACTAAT 2465
Db 1681 GCGCAGGTATGGGCTATAGGCGCTACACAGCGCGGAGGCTGATTCGCGAATAACA 1740
Qy 2466 AACACGCGCTGACGCTGTTTTCAGGCAATCACTAATAACGACTCACTCCCGTAAATATC 2525
Db 1741 TACCTCCTT-----ATTTTATAACAATAATGCTCCGCTGTGTGCA 1781
Qy 2526 CCGGAGTGCATTCATCTGCACCCCTAAAAACGAGGTTTATATGTCAGAAATAATAACAG 2585
Db 1782 GCGGAG-ACACTCACTCTGCACTCTGAAATCGAGGTCACATGTCGCTTAAATAACAAACAG 1840
Qy 2586 TGGTTGGCTTTGCCACTGCATCTGATATGGGATATATCGCCATCGCAGTATTTATGACT 2645
Db 1841 TGGTACGCTTACCCTCAACCTCATCTGGGATATGTCGCATCGCCGCTGTTATGAGG 1900
Qy 2646 GGTGATGTTTCGAACCTCGCATTTTATCGCACTATATTAATTCGTTAGGCTTCACACT 2705
Db 1901 GCGCAGCGCTTTGAGCTGGCTTTTCTGTCGCACATACATAAAGCACTCGGATTCACCCCG 1960
Qy 2706 GCGGAGGCTCTTTTGGCTTTTACGCTCTACGCGCTGGGGGCTGCCCTTTCCGCCCTGGGT 2765
Db 1961 GCGCAGGCTCTCTTTGGCTTTTACCTCTACGCTTGGCTGGCGCGCTGTCGCGCTGGGTC 2020
Qy 2766 TCTGGGTTAGTAGCGGAATCATACGCGCAAAACCATGCTGATTTGGTTTTGTCCTA 2825
Db 2021 TCCGGCGTGGTGGCGGAATCATTTACCCCACTGAAGGCATGCTGATTGGCTTCGTGCTA 2080
Qy 2826 TGGTGGTATTCATGTTCTGTTTCTGCTGTTGATTAGGACAGGCAACATATGGGTTA 2885
Db 2081 TGGTGTGATTTCCAGCTACTGTTCTCTGCTCTCGGCTCGGTCGCGCACTATCGCGTG 2140
Qy 2886 ATCCTGCTGTTTACGGGATCGTGGTTAGCGTATCCGCTATTTCTCTACTCATTTATC 2945
Db 2141 ATCCTGCTGTTTATGCGATCCGCGGCTGGCTTACCCTGCTTCTCTCTACTCTTTATC 2200
Qy 2946 GTCGTTATTTATCAATAGCGGTAGCGGAATAATTCAGTTCTGCTCTGGGTTGGTACTGG 3005
Db 2201 GTCGCCATTATCCACAGCTCGCGACAGTTCAGCTCCGCTCGGCTGGGCTGTTCTGG 2260

QY	3006	GGGGTATATTCAGTAGGGATCGGTGTGCTGGCAGGATATATATCCAGTATTCACGATACCG	3065
Db	2261	GGGGTCTACTCGGTGATTCGGCGTGTTCGGCAGCATATATCCGAGCTTTACCATTC	2320
QY	3066	ATTATGGGTGAATGGGAACCTTATGGCTTGGCAGCTGGCGTCTCTGCTTGGCGGTGTC	3125
Db	2321	CACATTTGGGGAATGGCACCCTGTGGCTGGCGTCTCTTTGGCCCGCCGCGGCATTT	2380
QY	3126	ATTGCCATGATCTCCCTTGGCTCATGTATAAAGCCCTGGACATATGATATTAATTAATCC	3185
Db	2381	ATTGGCTGGTTCGATGGCTATACCGAAGCCCGCCGATATGCGAAGCTTGACCAT	2440
QY	3186	CGTGAGAAATTTGCAAGATTAAGTCGGGAGTAACATTTACTTTATACCAACCCGCAATAT	3245
Db	2441	CGTGAATAATTCGCCAACTGGGGCGCGCGACGCTCTCTATACCAACCCGACAT	2500
QY	3246	TTTCTCTCCAGTATTCGCGCATATAAATACCTTATCTGTTATTCGGTTTTCGGTCA	3305
Db	2501	CTGTTCTCGAGCATTTGGCGCATTTATCAACACCGCTGCTGTTGTTTGGGTGATC	2560
QY	3306	ATGCCAATGATGTTTGGATGAACCTGGGATTCACACCTCTGAATGTTGCAGGTCTGG	3365
Db	2561	ATGCCGATGATGTTTGTGATGAGCTGGGATTTACCACCTCGGAATGGCTGCAGGTATG	2620
QY	3366	GGGGCATTTTCTTTACCACTATTTTCTCTAATATTTTGGGGGATTTGGCAGAAAA	3425
Db	2621	GGGGCATTTCTTCTACTACTATCTTTTTCGAACGTTTCTGGGGATTTGGCGGAGAA	2680
QY	3426	ATGGGCTGATCGCTTATTCGCTGTTTGTGCTCGGATGCGACATCAAGTTTA	3485
Db	2681	ATGGGCTGATGAAGTTTATCGCTGTTGCTGGCTGATCGGATGGCGTTCCTCCAGCCTG	2740
QY	3486	GGGTTTTACTACATCGCGCAATCTTTGGTTCACAACTACTGATGCAATATTCGCGCG	3545
Db	2741	GGGTTCTATTACTCGCCCGACACTTCGCGCATATTTTGGCCATGCGCTGGTACCGCT	2800
QY	3546	ATTGCTCTGGGAATTTTGTGCTGATTTGTGCGCATTTGCGCGCTGCTTCCCGGACATG	3605
Db	2801	ATTGCGCTGGGAATTTTGTGCGCGGTTTGTGCGCATGCGCGGCTGTTCCCGCGCTG	2860
QY	3606	GAACCAAAACAAAGGTGCTGCAATCTCGGTTTACAACCTCTCTCGCGGTATGCTAAC	3665
Db	2861	GAGCGAACCATAAAGCGCGGCCATTTGCGTCTATACTGTACGCGCGGTGTCGAAC	2920
QY	3666	TTCTCGGCTCGGCAATTCGCGTGGTGTATTACCGTGGTTTAGCACTATCGGTGTGTC	3725
Db	2921	TTTCTGGCGCGGCAATTCGCGTGGTGTACTGCTGCCCTACTCTCAGCACTATCGCGTGT	2980
QY	3726	ATTGCTATACAGCATTTGATCTATTGGCCTTTGCTTGGCCATTCATTCGCGTTGAG	3785
Db	2981	ATAGCCTATACCGCTTATATATCTGGCCTTCTTCCCTGCGCGTAAATCGCGTTGAA	3040
QY	3786	CAGCCAGGATTC-----GTTCTGCGCCAGTGACTGAGAAGGCAATGAATATCTCC	3836
Db	3041	CAGCCTGGCTTTACCAGCATCAGCATGCGAACCCTTTACGCCCAACCGCGCAGATCT	3100
QY	3837	TGAACCAAGCAACGATCAGGCACTCATCTCTCCCTCATGGAGAGGATGATTTACAT	3896
Db	3101	TAATAACCGCATCTGGTACAAACACAGCAATCGCGGTGCGCGCGGATTTTGGCT	3160
QY	3897	CAGGCAATAGTGCATTTGTTATCGAGATAAAGCTCTGCGCGCGTATATCATGTTTCAG	3956
Db	3161	TATGCGATAGTCACTTTG---ATCCAGATAAATCTCTGACCGGCAATGATCAGCTTACG	3218
QY	3957	CGGTACGACCATCGAT	3971
Db	3219	CCGTGCGCCATCGAT	3233
RESULT	4		
LOCUS	AE009731/c		
DEFINITION	Brucella mellitensis strain 10789 bp DNA linear BCT 28-DEC-2001		

the complete sequence.	
AE009731 AE008918	
AE009731.1 GI:17985189	
Brucella mellitensis.	
Brucella mellitensis	
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;	
Brucellaceae; Brucella.	
1 (bases 1 to 10789)	
DelVecchio,V.G., Kapatral,V., Redkar,R.J., Patra,G., Muje,C.,	
Los.T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,	
Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,	
Mazur,M., Goltsman,E., Selkov,E., Elzer,P.H., Hagius,S.,	
O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyrpides,N. and	
Overbeek,R.	
The genome sequence of the facultative intracellular pathogen	
Brucella mellitensis	
Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)	
11756688	
2 (bases 1 to 10789)	
DelVecchio,V.G., Redkar,R.J., Patra,G. and Muje,C.	
Direct Submission	
Submitted (13-NOV-2001) Institute of Molecular Biology and	
Medicine, University of Scranton, Scranton, PA 18510, USA	
3 (bases 1 to 10789)	
Elzer,P.H. and Hagius,S.	
Direct Submission	
Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag	
Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA	
4 (bases 1 to 10789)	
Kapatral,V., Los.T., Ivanova,N., Anderson,I., Bhattacharyya,A.,	
Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,	
Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Haselkorn,R.,	
Kyrpides,N. and Overbeek,R.	
Direct Submission	
Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell	
Park Drive, IL 60612, USA	
5 (bases 1 to 10789)	
Letesson,J.-J.	
Direct Submission	
Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,	
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,	
61 rue de Bruxelles, Namur 5000, Belgium	
6 (bases 1 to 10789)	
O'Callaghan,D.	
Direct Submission	
Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue	
Kennedy, Nimes 30900, France	
Location/Qualifiers	
1..10789	
/organism="Brucella mellitensis"	
/strain="16M"	
/db_xref="taxon:29459"	
/chromosome="II"	
105..581	
/gene="BMEII0977"	
105..581	
/gene="BMEII0977"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAL54219.1"	
/db_xref="GI:17985190"	
/translation="MLDARFTPPASRKRLSAPGLRTFLAIADLWKLDQOQLLVGLP	
SRSTVYNWKAVERHRTLDVDVLMRLSAVLGIHQALVLYPGGAAGRWLHTPNGA	
SLEFGQPLQLVASCTQDGLMAVRFLDAARGGLYMEPKALDRAPHPYHDEVDVFS"	
578..1243	
/gene="BMEII0978"	
578..1243	
/gene="BMEII0978"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	

QY 372 TTTCATGCCAATGCGAGCGCTTATATTGGCGGCCAGTGGCTGAAGGTGATCCAGATGTC 431
Db 3473 TTCCACGCCAATGCCGCTCTCTATATTGGCGGCCCTCTGGAGGCGCATCCGATGCA 3414
QY 432 TGGGATCGTGTGTAATCTGAATATAAATGCGGCGTTTCGGCTGTCTCCGTGCGAGTCCTG 491
Db 3413 TGGGATCGCATGTTGAACCTCAATATAATGCGGCTTTCGGTACCGTCCATCCGCTTCG 3354
QY 492 CGCATATGATTGCGCAGAGTGGCGGATATAAATTTTACCAGTTTCCATCGCGGCGTC 551
Db 3353 CGCATATGATGGAAGCAAGACTGCGACATCATGACCAAGCTCTGTTGCTGCGCTC 3294
QY 552 GTTCCGGTTATCGGAAACCGATCTACACCGGTCAAAATTTTCCGCTTCAGGCAATTCGTA 611
Db 3293 GTCCCGGTGATATGGAAACCGATTTATACGGGTCAAAATCCCGTGCAGGCTTTTCG 3234
QY 612 CACATACCGCGCGCGAGTTCCTCAATATGCGTGCSTGTGGTGCSTGCTGCCAGCA 671
Db 3233 CACACGGTTGCGCGTCAGGTGCGCGCATGGTATTTCGGCTTGGCGGTCTTTCGCGGGC 3174
QY 672 CCAGTAGTCACTGCCCTGCTTGATGACTGGCCAAAGCCAAATGGAAGGCCCTTGCCA 731
Db 3173 CCGGTCTCACTGCGCTCATGACGACTGCGCGCAAGCAAGCTCGATGAGCGCTGGCG 3114
QY 732 AATGTTAGCTGATGCAACCGATTTGAATGCGGGAATCAGTATTGTTATGTTGACCCGC 791
Db 3113 GCAGGACCGCTGATGAGCGGAAAGAGTTGCGGAATCCGTGATGTTTCATGCTGACACG 3054
QY 792 TCGAAAATGTCAACGTCGAGATTTAGTATCTCCCTGGCAGTGTGATCTGTAAGG 851
Db 3053 CCGCGCAATATCACCATCCGATCTGGTATTCTTCGCAAGTCTCGATCTCTAAATG 2994
QY 852 CGCAATCATGACAATAACCAAAACCGTTATTGGTTGATGTTGGATCAGGCAAGTCCG 911
Db 2993 GG----AGCGACCATGACCAACTATTATCTGGCGCTGATCTCGGCAACCGGAAGCCCG 2938
QY 912 CGCGGGATTTTGATCTCAACGGATCTCTGATATCCCATGCCACAGAAATATCAGAC 971
Db 2937 GCGCGGCTGTTCGACGCGGTGGCAACATGTTGATCGGCGGCGGATATGCCCAT 2878
QY 972 TAGCGGCGCAGCGGAAGCGGTGGAACAGTCCAGCAGGAGATCTGGCAGCGGTCTG 1031
Db 2877 ATGCGCGAGCGGCGGTATTGTCAGCAGTCCGACGATATATGCGAGCGCTTTG 2818
QY 1032 TTCATGTTATGTAATGCGCTCACTCTGCGACAGCTTTGTGCACAAAGTGTGGCAGGAT 1091
Db 2817 TGAAGTGTTCGGAAGTGTGAGGTTGCTGGCGTCTGATCCGCAAGCGGTTGCCGCAAT 2758
QY 1092 CGSTTTTGTATGCCACCTGTTCTGCTGCTGCTGATGATAAAACGGTATCCATGCTCTG 1151
Db 2757 CGGCTATGATGCCACCTGTTCTATTTGCTGCTGCTGGCGAGGCGCAACCGCTTGCGCT 2698
QY 1152 CAGCCCGGAAGGAGATGCAAGCAAAATATCATTTGTTGATGATGATCAGCGGCCACCGA 1211
Db 2697 TGGTCTCTCAAGATGCGGCGGCAACATCATCTGCTGATGATCATCGCGAGTGA 2638
QY 1212 ACAGCAGAGCAATCAATGCGCACTCACCATCCGCTGCTGAACCTAGCTCGGTGGTAAAT 1271
Db 2637 GCAGGACAGCGCTATCAATACCAACCAAGCGGATGCTGCTGGTATTGTCGGCGGCGCAT 2578
QY 1272 TTCGCTGAATGGAACACCGAAATCTCTGGCTGGAAGAAATATGCCAGATCTA 1331
Db 2577 TTCGCGGAATGGAACCTCCGAAATCTCTGGCTGGAAGAACACAAAGCGGAAACGTT 2518
QY 1332 CGAACCTGCGGACAAATTTTTCGATCTGGCGATTTCTGACCTGGCGGCTACCGTGA 1391
Db 2517 TGCTGCCGCTGGCAGTTTTTCGATCTGACCGATTTTCTCATATGGAATCTCTGGGCA 2458
QY 1392 TTTAGCGCTTCAGTACGATGTTTACCTGTAAATGGAGCTGGTGGCAATGAAATCG 1451
Db 2457 TCTTGGCGCTGGCGCTGACCGTAACCTGCAATGGACCTATCTCAGCCATGAAAGCG 2398
QY 1452 CTGGGATCCAGATTATTTCCGCAACCATTTGGCCTTCAGAGATTTAGCGGATGAAGATTTAT 1511

Db 2397 TTGGACGAGACCTATTTCCGCGCAGTTCGGCAACTCGCGGATGAAAGTTTCGT 2338
QY 1512 TCGCATTTGCTCATATTTTTCCTCCGCAACACACTTGTGGAAATGTTTAAACAGACA 1571
Db 2337 CGCATCTGTTACAGAGTTTCGGCGCGGTGGCGAAATCTTGG---CGGTTGAGCAACA 2281
QY 1572 AGCCGCGCAGAGATGGGATTACTCCCGGCACACCTGCTGTAGGTTAATGATGC 1631
Db 2280 GCGCGCAGCCGAACCTGGGCTTGAGGCGGCTACGGCAATTCGCGCGGCTGATCATGTC 2221
QY 1632 TCACGCTGTGGCATCGGTACGGTTGGCG-----TAGAAGTTGGAGCGCTGAACAA 1682
Db 2220 ACATCGCGCGGTATCGGCACGGTGGCGCGCGCGGTTCGGAAGGTTCGATATTTCGCG 2161
QY 1683 TCTCGGTATGTTTTCGGCACCTTCTTCATGCACATGGCATCTACCACTTCTCCCTCGTT 1742
Db 2160 TATGGCTATGTTTCGGCACATCGGCTTGACCATGACCACTGACACAGCTGTTTT 2101
QY 1743 TGTACCGGTGTCTGGGCGCTTATACAGTGGATGGTTTCCAGTCTGTGGTTAGTTGA 1802
Db 2100 CGTGGATGTTCTGGGCGCTTATTTCTCGGCATGGTTCGCGGCTGTGGTTGAACGA 2041
QY 1803 AGCGGGCAAGTGTGCGGAGCAGCTATTGACCACTACTTGTATTTCCATCCGCTGT 1862
Db 2040 AGCGGACAAATCGGCTGCGGTTGCGCTATTGATCATCTGATTCACATGCATCCCTTCGC 1981
QY 1863 TGAAGAAGCTCGCAAAATGGCACACAGTGTGAATCAGCCCTCCCGCTGCTGCTTCTGA 1922
Db 1980 GCGGAAGCGGAAGGACGCGCGGATCAAGCAAGGGCTGGCAGACAGCTCCTGCTGC 1921
QY 1923 TCGAATCTCGAAAAACCGCGCAACCATCAGATGCTGTGCGCCTGGCGAAAGGGTACA 1982
Db 1920 CGAAG---TTGAAGCTAGAGTGTCTGAAAAACCGGCATGATGCTTGGCGACATTC 1864
QY 1983 CTGTGTGCGGAATTTCTCGGAATCGCGCCCTTCGAGATCTTCATGCCAGAGCGGT 2042
Db 1863 TGTGCTGCGCGAATCTCTCGTAATCGTACCTTTCGCGGATCTGTGCGGCTGCGGT 1804
QY 2043 AATTTGTGCGCTTGGTATGAGCAGATCTGGATAATTTACTGCGCTTGTATATCGCTGG 2102
Db 1803 TATTCGCGGCTGGACCTTGATACAGGATGGACAGTCTGGCGGCACTTTATCTCGCGG 1744
QY 2103 ATTATCGGAATTTGTTATGCTGCGCCAAATTTCTGAGGCTCAACAGCAGCGAGGT 2162
Db 1743 GCTTTGCGGCTTGGCTATGTTGTCGCGAGATTTATCGAAGCGCAGCGGCCAAAGTAT 1684
QY 2163 AGTGACTAAAAATATCGTTATTAGCGCGGTGCGCGGAGCATCCACTGTCAGCAAAAT 2222
Db 1683 GTGACTGACATATTGTTGTTGCGGCGCGCGGCGCTCAAAATCTCGTGGCGGAGGT 1624
QY 2223 TCTGCGCATACCTGCGGTATTTCGGTCAATTACACGCAATGCTGCAACCGGTTTTTAT 2282
Db 1623 GCTCGCGATGCGACGCGCTCTGTTTACCGCTTCGACATCGCGGAGCGGTTCTGCT 1564
QY 2283 AGCTCGGCGCATTTGTTGCTGTCGCGGAAATATTGACACTTCTGTTGGCGAAGCGAT 2342
Db 1563 CGGCTCGGCGATTTAGGTGCGGTTGCGTCCGTTGTTATCCGATCTTGTGACGCTAT 1504
QY 2343 GCAACAATTCACCATGTGATAAATATTATTCGCGAAGACGCTATCAATCTCTTCA 2402
Db 1503 GCAAGTCATGTCGGAATTTAGGCGCAGCAACAGCGCGGAGCGCGGTCGCAAAATGGCA 1444
QY 2403 TCATCGTCGATATGAGCGCTATAAGCAGTTGACGATATCTGC 2444
Db 1443 TGACCAACCGCTTTGAGCGCTTCATGTTGCTTCAGGCGACTGC 1402

RESULT 5

AF045245/c

LOCUS

DEFINITION Klebsiella pneumoniae D-arabinitol transporter (dalT), and D-xylulose-kinase (dalK), D-arabinitol dehydrogenase (dald), and

5930 bp DNA linear BCT 30-AUG-2001

QY	3081	GGAACTTATGGCTGCACACTGGCGTTCTGCTTTGGTGGCGGTGTCATNTGCCATGATCTCC	3140
Db	1182	GGCACCCTGTGGCTGGCGTGGCTTCTACCGGGGGGTGATGCCCTGGTTTCC	1123
QY	3141	TTGCGTCATGTTAAACGCGCTGGACATATGCATAATTTAACTCCCGTGAAGTTTGA	3200
Db	1122	CTGCGCCATATTCAAACGCGGAGCATATGCAGAACTCTCACCACCGTGAAGTTTCC	1063
QY	3201	GAATTAACTGGCGCAGTAACCTTTACTTTATACCAACCGCAATATTTTCTCTCCAGTAT	3260
Db	1062	GAGCTGGAGCGCGCGAGCTGCTCTATACCAACCGCAATATTTCTGTCGACATG	1003
QY	3261	GTGCGCATTTAAATACCTTATCGTTATTCGGTTTTGCGGTCATTTATGCCAATGATGTTT	3320
Db	1002	GTGAGTGCATCAATACCTCTCGCTTGTTCGGCTTTGCGGTCATTTATGCCGATGATGTTT	943
QY	3321	GTGGATGAAGTGGGATTCACCACTCTGAATGTTTGCAGCTCTGGCGGCATTTTCTTT	3380
Db	942	GTGATGAGCTGGCTTTCAGCACTTCAGAAATGCTCCAGGTATGGGCACTCTTCTTCTTC	883
QY	3381	ACCACATTTTCTCTAATATTTTGGGGGATTTGGGCAGAAATAATGGGCTGGATCGCT	3440
Db	882	ACCACATTTTCTCTAACGTGCTGGGGATCCTCGGGGAAAACTGGGCTGGATGAAG	823
QY	3441	GTTATTCGCTGGTTGGTTGCGCTGGGATGGCAGCATCAAGTTTACGCTTTTACTACATG	3500
Db	822	GTGGTGGCTGGTTGCGGCTGCATCGGGATGGGCTCTCCAGGCTGGCGTTTACTACATA	763
QY	3501	CGCAATACTTTGGTCACAACTACTCGATGGCAATGATTCGGCGATTCCTCTGGGAAC	3560
Db	762	CCTCAGCACTTTGGCCATAGCTTTGGCATGGGCTGATCCCGCATCGCCCTCGGGATC	703
QY	3561	TTTGTGTCGCAATTTGCGCGATGGCGCTGTCTTCCCGGCATCGGAACAAACAAA	3620
Db	702	TTTGTGCGCGCTTCTGTCGCGCTGGCGCGCTGTCTTCCCGCGCTGCGAGCATAAA	643
QY	3621	GGTCTGCAATCTCGGTTTACAACTCTCTCGGGTATGTCCTAACTTCTCGCTCGCGCA	3680
Db	642	GTGCGGCAATTTTCGGCTCTATAACCTGTGCGGGGATGTGCAACTTCTCGCGCGGG	583
QY	3681	ATTGCGCGTGGTGTATTACCGTGGTTTACGATATCGGTGTGTCATTCGCTATACAGCA	3740
Db	582	ATTGCGCGTGGTGTGTCGCGCTTCTTCAGCACCATCGGGGTGTGATCGCCTATACCGCA	523
QY	3741	TTGATCTATTGCGCCTTTGCTCTTCCGCAATTCATTCGGGTTTGAGCAGCAGGATTC	3797
Db	522	TTGTACGTGGTGSCCTTTTCTCTCTCGCGCTTATTCGCGGTGGAGCAGCGGGGTTTC	466
RESULT 6	AE008252/c		
LOCUS	Agrobacterium tumefaciens str. C58	10295 bp	linear BCT 18-DEC-2001
DEFINITION	.187 of the complete sequence.		
ACCESSION	AE008252	AE007870	
VERSION	AE008252.1	GI:15158917	
KEYWORDS	Agrobacterium tumefaciens str. C58 (Cereon).		
SOURCE	Agrobacterium tumefaciens str. C58 (Cereon)		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
REFERENCE	1 (bases 1 to 10295)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 10295)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
COMMENT	Approximately 800 bp of telomeric sequence missing from the left		

FEATURES		end of the chromosome and 200 bp missing from the right end.
source		Location/Qualifiers
		1..10295
		/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
		/strain="C58"
		/db_xref="taxon:181661"
		complement(195..704)
		/gene="AGR_L_1070"
		complement(195..704)
		/gene="AGR_L_1070"
		/note="hypothetical protein"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1070p"
		/protein_id="AAK89112.1"
		/db_xref="GI:15158918"
		/translation="MAAEKRWGYPHINERPAHGRALMKSGEMSVTMLGHLFOLIL
		RPRHRVDVAFIAHEKVAERDRLRADAETADIDGGGGSDTMDMWLADLMV
		VCAINGCPFEHNGGEFRENEAYVIGMVHCSSCLGLHKTNSRTAKPFILOSVEVGY
		FSSVHCWGS"
		747..2297
		/gene="AGR_L_1071"
		747..2297
		/gene="AGR_L_1071"
		/note="PROLINE/BETAINE TRANSPORTER (PROLINE PORTER II) (PPII)"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1071p"
		/protein_id="AAK89113.1"
		/db_xref="GI:15158919"
		/translation="MVRHSRQQNLNQISQASPLSIDVTVIDDKVKAVTAAALGN
		AFWDFDVGVEYAVGVKFPFPAAPVQVAAATATFSPFLIRPLGGIFTCAMGDK
		FGROKVLSTIIIMAASTFCIGLPGYATIGWAPILLLCKLVGVSFGGYTGAAI
		FAEYAPDRKRGYLCGLDQSGIACFVGLAGLVLLSSAMGEAKFLDWGRIPFLAA
		PLGLGLYLRUAETPAFTERLQRAEEDRQSLKERPNVPAETAEIAREHFRSLAICIG
		MYLVNTVTYMLLTPTLYSKLTNYSEDHGVLLIIIVAWGMLFVQPAIGLSDRIGR
		KPFLAAGSTAILFLSLPAFLIASGVQVIFEGLLILAVALNCLIGINASTLPALFPA
		RIRYGALAFINISIVAGLTPTVTVAWLVETENTYMPAYLMAAVFGIVTTFLEKE
		TANRPLVGDPTNASSRKEAKSLLAQYQHIIEKSVEIDELAKLEEQKILQTKREGI
		VGRHPDLT"
		complement(2383..3195)
		/gene="AGR_L_1073"
		complement(2383..3195)
		/gene="AGR_L_1073"
		/note="carboxyl reductase"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1073p"
		/protein_id="AAK89114.1"
		/db_xref="GI:15158920"
		/translation="MRRVPFAPPALPDQSLGRYLWPNONNDSMOPSGKSVITGAGK
		GIGRACAOILMAARGAEVVAISRTQSDLSLKEIGRSIRVLDLADIAATRAAMAEAGP
		CDPLINAGINVLVSVDMSDAGYEAIVGINLRALVTCQEPARGVRANGGCAIVNI
		TSIAGHRGFDHLCYAAKAGLEGATVLAKELGPHGIRVNAVAPITITULEAEAWN
		DPEKSPQMVRHPLNRFPAEAEVAQSIALLSDSDSKMISGAVLPVDGGFLAV"
		complement(3158..4768)
		/gene="AGR_L_1075"
		complement(3158..4768)
		/gene="AGR_L_1075"
		/note="D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1075p"
		/protein_id="AAK89115.1"
		/db_xref="GI:15158921"
		/translation="MASVLVGVDTGTSARAGVFDVAGKLLATAKRPIISMHRDGGIA
		EQSSAEVWQAVCDSDSVSRAGIDPAEVTGIGFDTGCTSLVVRGPDVTLPVGAADHP
		ERDIIVMDHRAVEQAEIRINAGKAVLYVGGIRSPQETPKLLLSNRPDIYARAE
		HFEDLTDFLTWKASCALDRSACTVCTCKTYLAHENRWDAEYFTKIGLDLAPQGFRI
		GESVYHPGTALGNGLTEGAAKMGLVAGTVANAGLIDAHAGGVGTVAAGGDSRCLGY
		VGTSSCTWTMTTEPAFVFGVWGPYISAWPGAWLNEGQSAAGAAIDLYVLQHPAFA

```

gene
CDS
EAKALADKQKALPWLADRALSLAASAAVKLAEDFHMVPEF LGNRAPFADPHARA
I1AGYGETGVDSVLVYAGLGLGROI IETQARNGAPVETISVSGAGAHPLA
ROLLADTGLPVLTECEEPVLGSLGAMLGAAGCTY PDLMAAMPMSRIDSSAMPDPD
FOKTHQARYDAFLALQNAARARSASHS"
complement(4858..5589)
/gene="AGR_L_1076"
complement(4858..5589)
/gene="AGR_L_1076"
/notes="ribitol 2-dehydrogenase (EC 1.1.1.56) - Klebsiella
pneumoniae"
/codon_start=1
/transl_table=11
/product="AGR_L_1076p"
/protein_id="AAK89116.1"
/db_xref="GI:15158922"
/transl_table="MSESLQKIAVITGAASGIGLATTEALLEOGATVVMVDWNEKAL
NDLAAKIGERAIPQVNTLLDADSCNAMIPEILGKVQHDHIDILYNAGTYIGDGLTETTP
EADIMLNINLVANVKNQVAVPHMSERKTGDIIVTCSIAGHFTYWEPTVYSGSKWAI
TSFVQGMRRQMPHGVRVAQSPGVVPSYALLADWPEENLRKAKESGSLDASEVADAV
VYMLTRRTVTIRDMVLVPTNFDRV"
complement(5629..6612)
/gene="AGR_L_1078"
complement(5629..6612)
/gene="AGR_L_1078"
/notes="(AP001519) ribose ABC transporter (permease)"
/codon_start=1
/transl_table=11
/product="AGR_L_1078p"
/protein_id="AAK89117.1"
/db_xref="GI:15158923"
/transl_table="MMSLSLDANTGAAKTGTFSSAMLSPLALPLAGLIVVSTLMGL
ASDNFFSVNNIMNVLRQVSVVYILAVGMTFVILTGKGLDLSVGAVMALVGLTSLAGLMVN
TGLSPALTFAGVLGIGIGIFNCALVWAGMPAIIVTLATMGMARGLGLIYSGYVPV
SGIPSWLSPGVRGIVPVVIMVYIYAVAVLLORTAFGRHVYALGNELAARLS
GVKTRKVLAVYSGVTAFAALILFGLMSGPNAGVGFELDAIARVVLGCTAAG
GRGILGTGLAVLGLNLGLNGLMGINPLYQDVKIGIILLIYIGREWR"
complement(6596..8014)
/gene="AGR_L_1081"
complement(6596..8014)
/gene="AGR_L_1081"
/notes="1081"
/codon_start=1
/transl_table=11
/product="AGR_L_1081p"
/protein_id="AAK89118.1"
/db_xref="GI:15158924"
/transl_table="MHALLGNGAGKSTLMKILGILQPSSEGTIRIDGKORRFANYDE
AIAGVGVIFQEFSLIPLYNNAVENWFLAREIRGPERLLNKGAMKRAETMGRLLAVDQ
PLDPVFRSLVQAQQFVEIARALSLDARILVLDPTATLPSETEHLFNVNRELRRQV
VAIIFISHLDEIFEICDRITVLRDGLIGSCLTSEVNDRLVEMVGRRTTEASFPPK
POIDPSAAKVEVEBELQKGGPVSRFSLRKEILGFLAGLVGSGRTETVLAHLGAHSA
SRRIKLDGVTRLSGPDALMRGIGILLPESRKEGLITFSILQNTISLNLRYKRYKA
HWFILDLKLEHTTKAMAQOVKAQGPYARVDTLSSGNGOOKIVTARLNLHDMRVLIPT
EPTRGIDVAKAEIYSLMREPAARGHSIIIMISSELPVEIGMSDRVCFVRSRGIVATVE
GEDINSETIMNTTGRVHVA"
complement(8170..9081)
/gene="AGR_L_1084"
complement(8170..9081)
/gene="AGR_L_1084"
/notes="(AE005642) putative periplasmic ribose-binding
protein of ABC transport system"
/codon_start=1
/transl_table=11
/product="AGR_L_1084p"
/protein_id="AAK89119.1"
/db_xref="GI:15158925"
/transl_table="MKMKSFLMATAIAMAAPALAOEKLKIGMTFOELNNPVFVSM
QEALKEAASTGADVVYTDAGHDVAKQISDVDEMLQOKIDILLNPNDSAGIEAAVHA
AKAAGVTVAVNAFANGPVDTFVSGKNRDAGYKSKGLGDLGKGEVAILDGPVVPV
ILQVRECKAALAEFADIKLVDTONGQDRSVLGVVVENMIQSRPNLAGIFSVNDGGA
MGALAAIQGSKDKIKLTSVDGAPEAVKAIADGPFPIETTAQFPDQVRVGLAMALAK
WGARVPKPEVPIVWVDSKNAGFSW"
complement(9179..10096)

```

```

CDS
/gene="AGR_L_1086"
complement(9179..10096)
/gene="AGR_L_1086"
/notes="transcription regulator MtJR PA2337 (imported) -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_1086p"
/protein_id="AAK89120.1"
/db_xref="GI:15158926"
/transl_table="MRRRLMGTMQPDLELVHIRKGEFSAARHGHVFPFTRVRRHHY
PEYBIHLVATSGTFYIGDFVRFSPGOLIMTGNLNPWSEIPSEIDPVPTRSLVIO
FPESFIEDACTSMAEDVVRALLRRGILFDDDETSDRVPLRLEAAGLRLAL
FWEILDLVNAPESSVSLASVYELDTGIGDSGNRLSHLRHLETCQVEIDILADMV
GQSPSLKSRAPKRTHTGTTLVRYNRQLRIDLACOMLLADQDDVKVAEVCYDVGFNSLNF
NRHFLKLGMSPSRPTIFAAOKAAVAAE"
BASE COUNT 2108 a 3120 c 3036 g 2031 t
ORIGIN
Query Match 16.2%; Score 643.2; DB 1; Length 10295;
Best Local Similarity 55.2%; Pred. No. 8.8e-177;
Matches 1331; Conservative 0; Mismatches 1003; Indels 76; Gaps 1;
Qy 111 GTGCCCTCTATGAATACCTCCCTTAATGCAAAAGTTGCAGCTATCATCTGGCGCTGGGTCA 170
Db 5598 GAGAACTAAATGCTGTAATCGCTCAAGCAAGATCGCGGTCTATCACCAGCGCCGATCC 5539
Qy 171 GGTATTGGCCTCAATGTCRAAAAGCTGCTCGATCGAGGAGCAAGGTACTATTGATT 230
Db 5538 GGCATCGGGCTTGCACACAGACAGAGCGCTCTTGGAAACAGGCGCGCATGTCATGGTC 5479
Qy 231 GACCGGGAAGGCGCAAACTGCACAAAGATTGCTGCTAGTTAGGCGAAACCGCTACGCG 290
Db 5478 GACTGGAACGAGAAAGCCTTGAATGATCTGGCGCCAAAGCTCGGAGCGCCCATTCG 5419
Qy 291 CTGCAACTCGATCTCTCAATAATCAGCAAGTCGATACATCTCGGCGAGACATTTATCGAA 350
Db 5418 CAGGTGACCAACTGCTCGATCCGACAGCTGCACAGCGATGATCCGGAATCTCTGGGG 5359
Qy 351 CTGGCGGTGGCTGGATATTTTTCATGCCAATGCAGGCGCTTATATGGCGCCCGCATG 410
Db 5358 AAGTTCGATCATATCGACATCTCTTATTAATCGGCGACCTATATCGTGGTGTCTC 5299
Qy 411 GCTGAAGTGTATCCAGATGTTCTGGGATCGTGTGTTAAATCTCAATATAAATCGCGGTTT 470
Db 5298 ACCGAAACGACGCTTGAAGCCATCGACAAGATGCTGAACCTCAACGCTCAACCGGTATG 5239
Qy 471 CGCTGTGCTGCGAGTCTCTGCCGATATGATTGCGCAGAGTTCGGCGCATATAATTTTT 530
Db 5238 AAGAACGTGCAGGCGCTGCTGCCGCACATGTCGAGCGCAAGACCGGTGACATCATCGTC 5179
Qy 531 ACCAGTTCATCGCGGGGCTGCTCCGGTTATCTGGGAACCGATCTACACCGCTCCAAA 590
Db 5178 ACCTGCTGATCGCGGCCACTTTCCGACCTATTGGAGCCCGCTCTATTTCGGGCTCGAAA 5119
Qy 591 TTTCGCTTCAGGCAATCTGTACACACTACCCGCCGCCAGGTTTCTCAATATATGCGGTGCT 650
Db 5118 TGGCGCATCACAGCTTCGTGCGAGGCATGCGCCGCCAGATCATCCGCAACGCGTGC 5059
Qy 651 GTGGGTGGGTGCTGCCAGGACCATGATCTACTGCGCTTGTGATGATGATGATGATGATGAT 710
Db 5058 GTCCGCGAGGTCTCTCCGGGCCGCTGCTCGGACCTTCTCGGCGACTTCTGGCGGAGGAA 4999
Qy 711 AAAATGGAAGAGCCCTGCAAAATGTCCTGATGTCACCGCATTTGAAGTGGCGGAATCA 770
Db 4998 AATCTCCGAGGCCCAAGAAATCGGCGAGCTCTCATGATGCGAGGCAAGTCTGATGCC 4939
Qy 771 GTATTGTTTATGGTACCGCTTCGAAAAATGTCACCGTTCGAGATTTAGTATCTCTGCT 830
Db 4938 GTCTGTTATATGCTCAGCGCAAGCGCACCGCTCACCATCCGCGACATGCTGCTCTGCGG 4879
Qy 831 GGCAGTGTGATGCTTAAGGGCGCAATCATGACATAACCAAAACCG----- 877

```

Db 4878 ACCAACTTCGACCGCGTTTAAATAATCTCTGGAGCGCGGAAGAACCGGAAACCGGTTCCCG 4819
QY 878 ----- 877
Db 4818 CGCTTCAACCTGTTCAAGCACGCGATGATCCGCTCCGAGGAGTATCGCAAAATGGCTTCCT 4759
QY 878 ---TTATTGGTGTGATGTGGGATCAGGAGTGTCCGCGCGGGGATTTTGTGATCTCAAGC 934
Db 4758 ATCTTGTGGCGTCGATGTGGGACCGCTCGGCGCGCGGTGTCTTTGTGATGTCCGCG 4699
QY 935 GATCTCTGCTATCCCATGCCACAGAAAAATACAGACTACGCGCGCCAGCGGAACCGCG 994
Db 4698 GCAAGCTTCTCGCCACCGCCAGCGCCGATCAGCATGACACGGGAGGATGGCGGCATTG 4639
QY 995 TGGACAGTCCAGCCAGGAGATGTGCAGCGGCTGTGTTCAATGATTCGTAAATGCGCTCA 1054
Db 4638 CCGAGCAATCGAGCGCAGAGGTGTGCGAGCGGTTGTGATAGCGTTCGCGACAGGTTT 4579
QY 1055 CTCTGGCAGCGTTTGTGCACAAAGTGTGGCAGGCATTCGTTTGTGATGCCACCTGTTCTC 1114
Db 4578 CACGCGCGGTATCGATCCGCGGAGGTTACCGGCATTTGGCTTCGACGCGACCTGTTCTC 4519
QY 1115 TGGTGTACTGGATAAAACGGTGATCCATTCGCTCTCAGCCGCGAGGAGATGCAAGC 1174
Db 4518 TTGTGGTTCGCGACCCGGTGACGTGACCTCGCGGTAGTGTGCTCGCATCTCCCGAAC 4459
QY 1175 AAAATATCATTTGTGTGGATGATCACCGCGCCACCAAGCAGAGCAATCAATGCCA 1234
Db 4458 GTGATATCATCGTCTGATGGACACCGTGTCTCGACGAGCTGAACGCATCAACGCGC 4399
QY 1235 CTCACCATCGGTGTGAACCTAGCTCGGTGGTAAATTTGCGCTGAAATGGAACACCGA 1294
Db 4398 GCAAGCATGCGCTCTGAATATGTTGGCGCGCATTTTACCGGAAATCGACAGCCCA 4339
QY 1295 ARAATCTCTGGCTGAAGAAATATGCCAGAGATCTACGAACTGCGCGGACAAATTTTCG 1354
Db 4338 AGCTGTGTGGCTACGGAACAGCGCGGATATCTATGCCGCTGCCGAACACTTTCTCG 4279
QY 1355 ATCTGCGCATTTTCTGACCTGCGCGGCTACCGGTGATTTAGCGGCTTCAGTATGCACGT 1414
Db 4278 ACCTGACGGATTTTCTGACCTGGAAGGCTCCGGGCGCTTGATCGCTCCGCGTCGACCG 4219
QY 1415 TTACCTGTAAATGGAGCTGGTGGACATGAAATCGCTGGGATCCAGATTAATTTCCGCA 1474
Db 4218 TCACCTGCAATGGACCTATCTCGCCACGAAACCGCTGGGATCGGGAATATTTCCACCA 4159
QY 1475 CCATTGCGCTTCGAGAGTTAGCGGATGAAGATTTTATTCGATTGTCATCATATGTTT 1534
Db 4158 AGATCGGCTTGGCATCTGGCGGACGAGGCTTTCGCGCATCGGTGAAGCGTGTCC 4099
QY 1535 CTCGCCGACACCTTCTGCGTAGGCTTAATGATCCTCACGCTGTGTCATCGGTACCG 1654
Db 4038 TCGCAGGAACCGCTGCGCGCTGGGCTGATCGACCCCATCGCGCGGTGTGCGTACCG 3979
QY 1655 TTGGCTGAGAAGTGGAGCGCTGAACAATCTCCGCTATGTTTTCGGCACTTCTTCATGCA 1714
Db 3978 TGCTCGGGGTGTGATGATCAGATGCTCGGTTATGTTTTCGGCACCTCGCTCTGCA 3919
QY 1715 CCATGCGATCTACCACTTCTCCCTGTTTGTACCGGGTGTCTGGGCGGCTATTACAGTG 1774
Db 3918 CCATGACGACCACTGAGCGCGCTTGTGCGGGGTGTCTGGGGTCCCTATTATTCG 3859
QY 1775 CGATGTTCCAGTCTGTGGTTAGTTGAAGCGGGCAAGTGTGCGGGAGAGCTATTG 1834
Db 3858 CAATGGTTCTGCGCATGGCTCAATGAAGGGGCGGACGTCGCCGCGGTGCGGCGCATG 3799
QY 1835 ACCAGCTACTTGAATTTCCATCGCGGTGTTGAAGAACTCGCGAAATGGCACACGTTGA 1894
Db 3798 ATTACCTTGTGCAATGTCACCGCGCTTTCGCGAGCAAAAGCGCTTCCGACAGGACG 3739

QY 1895 ATCAGCCCTCCCGCTGCTGCTTGTGATCGAATCTCTGAAAAACGGGCAACCATCAG 1954
Db 3738 GCAAGCATTTGCCAGTCTGGCTTGCCGACGCGCTCAGCCTCGCGCATCGCTCCG 3679
QY 1955 ATGCTGTCCCGCTGGGAAAGGCTACAGTGTGCGGGAATTTCTCGGAATTCGCGCC 2014
Db 3678 CGCAGTGAAGCTTGCCGAAGATTTTATATGTCGCGGAATTTCTTGGCAATTCGCGCG 3619
QY 2015 CTTTCAGATCTCATGTCAGACGCGTAATTTGTGGCTGTGGATATGGAGCAGATCTGG 2074
Db 3618 CTTTCGCCATCCCATGCTCGGCCCATATTGCCCGTTATGGCATGGAGACGGCGTGG 3559
QY 2075 ATAATTTACTCGCTTGTATATCGTGGATTTATCGGAAATTTGGTTATGTTCTCGCCAAA 2134
Db 3558 ATTCCCTCTCGCGCTCTATGTGCGGGCTTCTCGCCTCGGTTACGGGCTTCGCCAGA 3499
QY 2135 TTCTCGACCGCTCAACAGCGCAGGAGTGTGAGTAAAATATTCGTTATTAGCGGGTG 2194
Db 3498 TCATCGAAAAACAGCCCGCAACGCGCGGTGGAACCATCAGCGTCAAGTGGCGGAG 3439
QY 2195 CGGGCAGCATCCACTGGTACGACAAATTTCTGGCAGATACCTCGCGTATTTCGGTCAATTA 2254
Db 3438 CAGGCGCTCATCGCTTGGCAGCAGTTGCTGGCAGATGGACCGGCTCCCGGTTGAAC 3379
QY 2255 CCACGCAATGCTCGCAACCGGTTTTATTAGGCTCGGCCATTTCTTGGTGTCTCGCCGAA 2314
Db 3378 TGACCGAATCGGAGGCGGCTGCTGTTGGATCTGCCATGCTCGCGCGGTTGCGCGCTG 3319
QY 2315 ATATTGACCTTCTGTTGGGGAAGCATGCAACAAATTCACCCATGTGGATAAATATTAT 2374
Db 3318 GAACCTATTCGGACCTGATGGCTGCCATCGCGCGCATGTCGCTATCGACAGCAGTGCA 3259
QY 2375 ATCCGCAAGACGCTATCAATCTCTTCATCATCGTATGAGGCTATAACGACGTTGC 2434
Db 3258 TGCCGATCCGATTTTCAGAGATTTCACAGCGGCTTACGACGCTTCTTGGCGCTGC 3199
QY 2435 AGCATACTGC 2444
Db 3198 AGAATCGCGC 3189

RESULT 7

AE009360

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AE009360 11437 bp DNA linear BCT 20-DEC-2001
Agrobacterium tumefaciens str. C58 linear chromosome, section 130
of 187 of the complete sequence.

AE009360 AE008689
AE009360.1 GI:17742779

Agrobacterium tumefaciens str. C58 (U. Washington).
Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

1 (bases 1 to 11437)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendinning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M.,
McLellan, E., Palmieri, A., Raymond, C., Rouse, G., Paulsen, I.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.

The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)

21608550
11743193
2 (bases 1 to 11437)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.

TITLE Direct Submission
JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

FEATURES
source

location/Qualifiers
1. .11437
/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
/strain="C58"
/db_xref="taxon:180835"
complement(68..1297)
/gene="Atu4316"
complement(68..1297)
/gene="Atu4316"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein"
/protein_id="AAL45110.1"
/db_xref="GI:17742780"
/translation="MNSPYREIFRPGAKGFSAGFFARLPPIAMAPIGIVAMLSQTHG EYWLAVGATFALTNAAPQISRLVDRMQSRVLPATIIISVIAFALITATNORW PTWLFUSAFLAAMPISIPAMRARTEIFRDRPELNTAFESAQDELIVYIAGASLS VGLAVFPFAGMMISTTFLAGTFAFLQRTSTPKVPVPSGYSRQSAIFLRPVQII TSLIFVGSFATAEVSVAITKELQPGAAISLVIGVYAGISGVVGLTGAIRPKMPL ORLLAVLAVLALLPLLVANSSVALLAFVLSGVAISPVFTTSGLTERRVPESM LTGVTWMTGIGIGMAEGAFISGWIDNPGDNGFWVSVASLTITVAIIGLQGRSLS GTRDSCASAVOQPAE"
1441..2013
/gene="Atu4317"
1441..2013
/gene="Atu4317"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, TetR family"
/db_xref="GI:17742781"
/translation="MTAETRAKLLAAGRKAFSGISGYAEASMDDETAGAGLTGALYHH FGDKGLLMVAVAEIDAEMTKRLCDISSAETRWEGVEENGVYRNALPEFIORNF RDGPVILGDPSPRPNANGCISAVARSLSALKKEETIGIDPEACARFINAASSTAQM IANSDDPETTSKRAVTSFRTFLEGLRIRKG"
2261..3313
/gene="Atu4318"
2261..3313
/gene="Atu4318"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="xylitol dehydrogenase"
/protein_id="AAL45112.1"
/db_xref="GI:17742782"
/translation="MKALVLEEKGLSLRDFDIPGGAGSGELGPKDVRIRTHTVGICG SDVHYTHGKIGHFVFNAPVNLGHEASCTVIEGSDVTHLKGIDRCVMEGIPDPTSR ASKGLIYNVDPAVRFWATPIHGLCTPIPEVIHPAAFTYKLPDNVSFAGANVEFPAIGM QALRIYQPDIAVYTAGPIGMVVALAGGCAKVIIVADQPKLDIIAAYDGLIE TIRIRNLAEAYSAATDGGWCDDIVFECGAPAILGMMAKLARPGGAIVLVGMPDVF PVDIVGLOAKELVETVFRVANYVDRAVALIASGKVDLKLISATIPFEDSIAGFDRA VEARETDVKLIQVMPQ"
3670..4569
/gene="Atu4319"
3670..4569

/gene="Atu4319"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, AraC family"
/protein_id="AAL45113.1"
/db_xref="GI:17742783"
/translation="MGTMQPOLELVHIRKGSFAAHRHGYPRTVRWHYHPEYEIH LVATSTGYIGDFVGRFSPGOLIMTPQNWISIEPDEIVPTSLVLTORPESFI EDACTSAEMDVVRLALDRSRGILFDDETSDVRPLVLRLEAGQTRCALFWELID LLVNPSPSEVLASLSYELDTIGDSCINRALSILRHLTEQVEIEDLADNVGSGPSA FSRFRKHTTGTTLVRYNRQLRIDLACOMLLAQDQVKVAECYDVGFSNLSNFRHFLK LKGMSPSRFRTIFAAQAAVMAE"
4667..5578
/gene="rbsB"
/note="Atu4320"
4667..5578
/gene="rbsB"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate binding protein [ribose]"
/protein_id="AAL45114.1"
/db_xref="GI:17742784"
/translation="MKMSKFLMATAAIAAMAAPALAEKLIKGMTFOELNPNPVFSM OEALKEAASIGADVVTVDAGHDAKQISDVEDMLQOKIDILLNPTDSAGIEAAVHA AKAGVIVVAVDANANGPVDTFVSGSKNRDAGYKSKHLGDLGKGGEVALIDGIPVVP ILQRVECAALAEFADIKLVDTQNGRQDSVAGLVVENMIQSRNPAGIFSVNDGGA MGLAAIQSGSKDILKTSVDGAPEAVKAIADGGFFIETTAQFPDQVRVGLAMALAQK WGARVPVEKPIDVWVVDKSNAGEFSW"
5650..7152
/gene="rbsA"
/note="Atu4321"
5650..7152
/gene="rbsA"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase protein [ribose]"
/protein_id="AAL45115.1"
/db_xref="GI:17742785"
/translation="MLENLARKSFGKIEVLHGVDEARAGEVHALLGECAGKSTLM KILCGILOPSEGTIRIDGKDRRANYDEATAAGVGVIFQFESLIPYLNVENMFLARE IRPLRLNKGAMKRAEIMGRUAVDVPDVPVHRLSVAAQQQFEIATKALSADARIL VLDEPTATLPSETEHLFNVNRELRRGVAIIFISHHLDEIFEICDRITVLRDGLIG SCITSEVNDRLVEMVMYGRRIEASFPKPQIDPSAAKVIIEVEELQKKGVPVRSFLR KGETLGPAGLVGSGRTETVLAMLCASRRKIKLDGVDVTLRSGPDDALMRGIGLLPE SRKEGLITSFSLQNTSLNRYRYRAHWFLLDKLEHTTKAMAQOVVKAQGPYAR VDTLSGNGQKIVIARWLNHDMRVLIFFDEPTRGIDVGAKAEIISLMREFARGHSIIM ISSLEPEVIGMSDRVCVFRSGGIVATVEGEDINSETIMTNATTGRVERHA"
7136..8119
/gene="rbsC"
/note="Atu4322"
7136..8119
/gene="rbsC"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein [ribose]"
/protein_id="AAL45116.1"
/db_xref="GI:17742786"
/translation="MNSLSDATGAATGGSFSAMLRSPALPLAGLIWVSTLMGL ASNFFSVNNIMNVROVSVVGLVAGMTFVILTGGDLSVGNALVGLTSGSLMVN TGLSPALPAGLFIGLIGICIGFNGALVANGMPAIIVTLATGMARGLGLIYSGYPV SGPSWISWFGVGVGVIPVPIIMVVIYAVWVLLQRTAFGRHVYALAGNELARLS GVKTRRVKLAVYGISGVTAAFAALILTGRLSMQPNAGVGFELDAIAVVLGGTAIAG

QY 1775 CGATGGTCCAGGCTGTGGTGTAGTTGAAGGCGGCAAGTGTCTCGGAGACGACTATTG 1834
 Db 9890 CAATGGTTCCTGGCGATCGGCTCAATGAAGGCGGACGTCGCCGCCGCGGCGCATCG 9949
 QY 1835 ACCAGCTACTTGAATTCATCCGCTGTGAAGAGCTCGCAAAATGSCACAACTGTGA 1894
 Db 9950 ATTACTTGTGCGAGTTCACCGGCTTTCGCGAGGCAAAAGCGCTTCGCCAAGGACG 10009
 QY 1895 ATCAGCCCTCCCGCTGTGGTGTGATCGAATCTCGAAAAACGCGCAACCATCAG 1954
 Db 10010 GCAAGCATTCGCACTGTGCTGCGAGCGGCGCTCAGCTCCCGCATCCGCTTCG 10069
 QY 1955 ATGCTGTCCCTCGGCAAGGCTACACGTGGTCCGGAATTTCTCGGAATCGCGCC 2014
 Db 10070 CGCGAGTGAAGCTTCGCGAAGATTTTCATATGATGTCGCGGAATTCCTGCAATCGCG 10129
 QY 2015 CTTTCGAGATCTCATCCAGCGGTAATTTGTCGCTGGTATGAGCGAGATCTGG 2074
 Db 10130 CTTTCGCGATCCCATCTCGCGCATATTCGCGGTATGGCATGGAGAGCGGCGTGG 10189
 QY 2075 ATAAATTTACTCGCTTGTATATCGTGGATTATCGGAATTTGTTATGCTGCGCCAAA 2134
 Db 10190 ATTCCCTCGTCCGCTCTATGTGGCGGCTTCTCGGCTCGGTTACGGGCTTCGCCACA 10249
 QY 2135 TTCGAGCTCAACACGCGAGGAGTAGTGATTAATAATATCGTTATAGCGCGGTG 2194
 Db 10250 TCATCGAACACAGCGCGCAACGCGCGCGGTGGAACCATCATCGCTCAGTCAGCGGAG 10309
 QY 2195 CGGCGACATCCACTGTACGACAAATTCGGCAGATACCTGCGGTATTCGGTCAITTA 2254
 Db 10310 CAGGCGCTCATCCGCTTCCAGGAGTGTGCTGGCAGATGCGACCGGCTCCCGGTTGAAC 10369
 QY 2255 CCACGCAATTCGCAACCGGTTTATTAGGCTCGGCAATTTCTGCTGTCGCGCGAA 2314
 Db 10370 TGACCGAATGCGAGCGGCTGCTGCTGGATCTGCCATGCTCGCGCGCTGCGCGTG 10429
 QY 2315 ATATTGACCTCTCTTGGGAGAGGATGCAACAAATCACCCATGTTGATTAATATT 2374
 Db 10430 GAACCTATCCGAGCTGATGCTGCATGCGCGGATGTCGGTATCGACAGCAGTGA 10489
 QY 2375 ATCCGCAAGACGCTATCAATCTTCATCATCTGATGCGATGAGGCTTATAGCAGTTGC 2434
 Db 10490 TGCCCGATCCCGATTTTCAGAGATTCACAGGCGGCTTACAGCGGCTTTCGCGCTGC 10549
 QY 2435 AGCATACTGC 2444
 Db 10550 AGAATGCGGC 10559

RESULT 8
 KPNRBTDK
 LOCUS KPNRBTDK 851 bp DNA linear BCT 26-APR-1993
 DEFINITION K.aerogenes ribitol dehydrogenase (rbdD and rbtK) genes, complete cds and 5' end respectively.
 ACCESSION M25606
 VERSION M25606.1 GI:149318
 KEYWORDS ribitol dehydrogenase.
 SOURCE K.aerogenes DNA, clones lambda-[prbt, prbt, dal] and pJCW1.
 ORGANISM Klebsiella aerogenes
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS Loviny, T., Norton, P.M. and Hartley, B.S.
 TITLE Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural gene
 JOURNAL Biochem. J. 230 (3), 579-585 (1985)
 MEDLINE 86050424
 PUBMED 2933028
 FEATURES
 Location/Qualifiers
 1. 851
 /organism="Klebsiella aerogenes"
 /db_xref="taxon:28451"
 28. .777
 CDS

/note="ribitol dehydrogenase rbtD"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25140.1"
 /db_xref="GI:149318"
 /translation="MKHSVSMNTSLSKVAATGAASGIGLECAARTLLGAGAKVLLI
 DREKLNKLVAEIGENAFALQMDQVNDLQGLTQLTGRDIDPHANAGAYIG
 PVAEGDPDNDVRLHINAAFRVSRVPLHIAQKSGDIIFTSSIAGVVPEPVY
 TASFVAQAFVHTTTRVOAQQYGRVAVLPVPTALLDDWPKAKMDEALANGSLMQP
 IEAVESVLEFVHTRSKNVTVRDIVILPNSVDL"
 804. .851
 /note="ribitol dehydrogenase rbtK"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25141.1"
 /db_xref="GI:149320"
 /translation="MHNDQNIIGVDVSGS"
 BASE COUNT 172 a 248 c 264 g 167 t
 ORIGIN 226 bp upstream of SalI site.
 Query Match 11.8%; Score 468.6; DB 1; Length 851;
 Best Local Similarity 75.9%; Pred. No. 1.3e-125;
 Matches 579; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
 QY 86 AAAGGATTATATGATGAATCACTCTGTGCCCTCTATGAATACTCCCTTTAATGGCAAAGT 145
 Db 15 AAAGGATTATAAATGAAGCACTCTGTCTCTCTATGAATACTCTCCCTCAGCGGTAAGT 74
 QY 146 TGCACCTATCACTGCGCTGCGTCAAGTATTGGCTTCAATGTGCAAAACCTGCTCGA 205
 Db 75 CGCCCGCATCACCGCGCGGCTCGGTTATCGGCTCGAGTCGCGAGGACCTCGTTGG 134
 QY 206 TGCAGGACAAAGTAGTATTGATTGACCGGAAGGACAACTGCACAAAGATTGTCG 265
 Db 135 GGCTGGCGAAAGTGTACTGATCGACCGCAAGGCAAGAGCTCAACAAACTGTCG 194
 QY 266 TGAAGTAGGCGAAACGCGTACGCTGCAACTGCACTCTCTTCAATAATCAGCAAGTCGA 325
 Db 195 CGAACTTGGCGAAACGCTTCGCCCTTCAGGTTCGACCTGATGACGAGGACCAAGTCA 254
 QY 326 TAACATGCTGGCGGACATATTCGAACTGCGGGTGGGTTGGATATTTTCATGCCAATCG 385
 Db 255 TAACCTACTGCAAGGCAATTTTTCAGCTTACCGCGGCTCTCGATATCTTCCAGCGCAACG 314
 QY 386 AGGCGCTTATATTGCGGCCAGTGGCTGAAGGTGATCCAGATGCTCGGATCGTGTGT 445
 Db 315 CGGCGCTTATATCGCGGGCGGTCGCCAGGCGGACCGGACGCTGGGACCGCGTGT 374
 QY 446 AAATCTGAATATAATGCGGCTTTTCGCTGTGTCGCTGCACTCTGCGGCTATGATTCG 505
 Db 375 GCACCTTAATATCAACGCGGCTTCGCTGCTGCGCAGCGCTCTGCGCCTCTGATCGC 434
 QY 506 GCAGAGTGGCGGATATAATTTTACAGTTCCATCGCGGGCGTCTGCGGTTATCTG 565
 Db 435 GCAAAATTCGCGGGATATATCTTCCAGCTGATTCGCGGGCGTGGTGGCGGTTGATCG 494
 QY 566 GGAACCGATCTACACCGGCTCCAAATTTCCGCTTCAGGCAATTCGTACACACTACCGCG 625
 Db 495 GGAGCTGTCTATACCGGCTCGAAATTCGCGTTCAGGCAATTTGTGCACACCACTGTCG 554
 QY 626 CCAGCTTTCTCAATATGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 685
 Db 555 CCAGTGGCGGATGATGCGTGGCGCTCGCGGCTGATTTTCCGCGGCGCGGCTGGTGGTGG 614
 QY 686 CTGCTTCTGATGCTGGCAAAAGCCAAATGGAAGAGCCCTGGCAAAATGGTAGCTGAT 745
 Db 615 TCTGCTGACGACTGGCCAAAGCCAAATGGAAGAGCCCTGGCAAAATGGTAGCTGAT 674
 QY 746 GCAACCGATTGAAGTGGCGGAATCAGTATTGTTTATGTTGACCGCTCCAAAAATGTCAC 805
 Db 675 GCAGCGGATTAAGTGGCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 734
 QY 806 GGTGCGAGATTAGTGTATGATCTGCTGCTGCGGAGTGTGATCTGTAA 848

```
||||| || | ||||||||| ||| || ||||| ||
Db 735 CGTGCAGCATGTGATCTGCCGACACGCGGTGATCTCTGA 777

RESULT 9
AE013623/c
LOCUS
DEFINITION
ACCESSION AE013623 AE009952
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
1 (bases 1 to 10201)
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
Genome Sequence of Yersinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
12142430
2 (bases 1 to 10201)
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,
Blattner, F. R. and Perry, R. D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. .10201
Location/Qualifiers
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
complement(126. .731)
/gene="y0230"
complement(126. .731)
/gene="y0230"
/function="unknown"
/note="residues 14 to 194 of 201 are 44.75 pct identical
to residues 14 to 194 of 200 from GenPept :
>emb|CAC95734.1| (AL596165) similar to putative
sugar-phosphate isomerase [Listeria innocua]"
/codon_start=1
/product="hypothetical"
/protein_id="AA083824.1"
/db_xref="GI:21956906"
/translation="MSQNKWLAASDAMIYSRELAALKENVDQVWLQVLDALAECR
GKIAYGVGTSGIAARKVAHMLACVEOPAIYNATDAAGDGLFGAODIILISRG
NSDELTRLTLQKQVKTIVTENEQSAIAQVSAVLKTHYKQIEDPLNMLATTISV
LVLAFDAICACLMARTGSKETLLAVHPGGDVGWVLRKQ"
complement(728. .2380)
/gene="y0231"
complement(728. .2380)
/gene="y0231"
/function="enzyme"
/note="residues 8 to 548 of 550 are 50.44 pct identical to
residues 13 to 566 of 569 from GenPept : >emb|CA81024.1|
(AL161576) putative protein [Arabidopsis thaliana]"
/codon_start=1
/transl_table=11
/product="putative sugar kinase"
/protein_id="AA083825.1"
/db_xref="GI:21956907"
/translation="WRDKTMSYFTFGVGTGSRAGVFDIGRMVQASREITWFKP
KADFVEQSSENTWQACNVRDAVNOADINPQVGLGFDATCSLVLDLKEGNPLTVS
PSGRNEQWLVWMDHRAITQAEINATKHPVLEFVGGVISPEMQTPKLLWLKQHPNT
WSNVGLDFDLPFLTWATKDETRSLCTVCKWYTLGHEDRDPSFKVLVGLADLLDN

NAAKTGATVKPMGAPLGHLSQRAASEMGLIFGTAVSVIIDAHAAGTIIILGASGVGTG
ENANFDRRIALIGTSTAHHMARSAAHIFSGVPIYSALPEYMLNEGQATGALI
DHIIOHPCYPALLQAKNKGETIYEALNIIILQWAGEPIEAFIETNDHMLPYDTHGN
RSPRANPLTGIITGLTKLSTPEDNALRYLTAOLALGTRHIIETMNGNYNDTMM
ASGGCTKNPIFVQEHANATGCAMLLPESEAMLLGSAMGTVAAGVFESLPAMAAMS
RIGKTVTQTNKIKAYYDKYRVFQHYHDHRYQALMQEGA"
complement(2384. .3337)
/gene="y0232"
complement(2384. .3337)
/gene="y0232"
/function="transport; transport of small molecules;
carbohydrates, organic acids, alcohols"
/note="residues 20 to 310 of 317 are 58.76 pct identical
to residues 16 to 306 of 318 from GenPept :
>gb|AAG54673.1|AE005211.8 (AE005211) putative permease
component of transport system, probably ribose specific
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="putative permease of ABC transporter"
/protein_id="AA083826.1"
/db_xref="GI:21956908"
/translation="MTWLNRTIPDDRIIRLQLLILIAVLMVFSLTIGORFFSLGNFOS
MSSOLPILGMLALGMLTMTGGINLSIIAGACSLVMAAIIIVSHPGQAPELLALV
AGLLVAIGTLNGVLISVIGVSPILATLGTMTLITGLNILLSNGDVISGFPPIQYI
GSGDIVGIPVAMILFLLVAAGLWVLEHTFLGRSIIYLVGSEQATRFSGVNRHVOIA
VYLSALLGWAAIILMAKFNKAGYGESYLLVTLIASVLGIPNDGGFGRIIGLIL
ALVYQLLESGLNLGVSYSYLTWALMGVLLIFALQNRKA"
complement(3339. .4328)
/gene="y0233"
complement(3339. .4328)
/gene="y0233"
/function="transport; transport of small molecules;
carbohydrates, organic acids, alcohols"
/note="residues 6 to 321 of 329 are 57.59 pct identical to
residues 1 to 316 of 323 from GenPept :
>gb|AAG54672.1|AE005211.7 (AE005211) putative permease
component of transport system, probably ribose specific
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="putative permease of ABC transporter"
/protein_id="AA083827.1"
/db_xref="GI:21956909"
/translation="MDKLRQLTGHHEFYLGVLGVLALLGLSVQSPFLTGLNLTDV
ATSYAILGILACGLFVLIAGDIDISFPATVIAQVMAASVITQSGSPPLALVMAA
VGLLGLINGLLVYLVKPAIITATLNLFYLLVLTGTFYFGFDPWFMTGINWF
VSGESGDYDGLTLLPLCLAGTIIFTGVMMVTRLCQRFAMGSKNDAASRLGINFR
LHLYVGYMGIAGVAAYVQAOISQVAPNSLMGFELTVLAAVVLGSGTSGRGSIT
GTVLGMLLAFQLNGLTLLSVSSYHVTFSVGIILVSIISTTANNEKRLRHH"
complement(4321. .5817)
/gene="y0234"
complement(4321. .5817)
/gene="y0234"
/function="transport; transport of small molecules;
carbohydrates, organic acids, alcohols"
/note="residues 116 to 498 of 498 are 61.19 pct identical
to residues 9 to 392 of 392 from GenPept :
>gb|AAG54671.1|AE005211.6 (AE005211) putative ATP-binding
component of transport system, probably ribose specific
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="putative ATP-binding protein of ABC transporter"
/protein_id="AA083828.1"
/db_xref="GI:21956910"
/translation="MTQATAFITLENISKRFPGVIALDGVNLTLNKGEVHCLAGNQC
GKSTIKVISGYQPEKASILIDGKLHLQPLTSFYGVQVIYDLSLFPNLTVAE
NTAVHRYLPGGDFVWKRSMRERAAAMQVGTDLDDKKVQLSTADQLVAICRAI
ADARALVMDPEASTLSQEVKLLNVRLDKSQGICVVFVSHRDEVMEVADICRVM
RDGKVGWQASSELDSHELAFALMTGORTYRPLPLAAKAAAPLLETIRKLSRGEQFNI
DLTLHQGEIVSITGLLAGRTCLSLSGMTQPESEIFVAGEYFVRFRNRDAIRHGI
GTVSEDRLTQGLIMEQSIYDNTIVSVFDQLTRSGLLDHSKAAALNKLVLGNTKVS
```

DTALPVKTLSSGNQRIATAIKWATOPRILILDSPTVGVDIANKKEGYHIAKALAEQG
 MAVLMICDEIPEAYINSHRVLVLRKGLIAEFYPHQCTEQOIAEVWNG"
 complement(6021..7079)
 /gene="y0235"
 CDS
 complement(6021..7079)
 /gene="y0235"
 /function="transport; transport of small molecules;
 carbohydrates, organic acids, alcohols"
 /note="residues 25 to 350 of 352 are 53.37 pct identical
 to residues 5 to 327 of 328 from GenPept :
 >qlAAG54669.1(AE005211.4 (AE005211) putative periplasmic
 binding protein, probable substrate ribose [Escherichia
 coli O157:H7 EDL933])"
 /codon_start=1
 /transl_table=11
 /product="putative periplasmic binding protein of ABC
 transporter"
 /protein_id="AAM83829.1"
 /db_xref="GI:21956911"
 /translation="MQHNKISYLATVPLPERKQEGIAKMFNLALLNVCISACMLFTT
 QTLAAEKHHEIAVVAKVGIPIWFTRMKEVGYNEAAKLNVAAYOVGPATPPPAQOVKI
 EDLIAKNVDALIVPNDKAKVLEPVLLKKAQEGIVLTHESPDORIGOWDETIDSEKY
 AQAMDELAKMNGKGYAIYVGSLLVPLHNADYAIKYQKEYPENFEVTPRLPVA
 ENDKSTETDLMLKTYPMQKGIIGFSGSLPIGAGQAVAKRKRKDOIAVVGIAWPAQA
 APYLMRGDIKKALLMDPKDAGFAVVEIANQLNGQKVTEDLTIDGLKADVDVDSKNGVI
 RFNKILEVTKDNAKTLGF"
 7723..8724
 /gene="ddg"
 /note="y0236"
 7723..8724
 /gene="ddg"
 /function="putative factor"
 /note="residues 25 to 333 of 333 are 66.66 pct identical
 to residues 20 to 327 of 328 from E. coli K12 : B2378;
 residues 28 to 333 of 333 are 66.66 pct identical to
 residues 1 to 305 of 306 from GenPept : >emb|CAD07638.1|
 (A1627274) putative acyltransferase [Salmonella enterica
 subsp. enterica serovar Typhi]"
 /codon_start=1
 /transl_table=11
 /product="putative heat shock protein"
 /protein_id="AAM83830.1"
 /db_xref="GI:21956912"
 /translation="MSEDMDYRYRMSLCVTHSHLTHNDRVHMIPKQFHISLLHPRYW
 LTFWGLVLELLVQLPYPLLNKLGVLGRTSMRFLKRRVSIARRNLELCFDMDKQVL
 EOTIIGFESLIGLLETHGMAWFSDFRIORFVSGLNKLKRAQCKRGVLVIGVHF
 MSLELGRVWGQCPQNMAMYRPHNNKVMELVOTWGRMSKNMLDRDLRLMKRPAKQ
 GEAWFAPDDQYGRGSVFAPLFAVEHAATSSTGFMLARLAKPALLPLVLRLMKRKGK
 YDLTIQALSDYPIPDIDEIAAASYNKVKIEIKMRAPEQYLWLHRRFRPTRPIGEPSLY"
 8768..10063
 /gene="y0237"
 8768..10063
 /gene="y0237"
 /function="unknown"
 /note="residues 186 to 247 of 431 are 26.15 pct identical
 to residues 515 to 579 of 865 from GenPept :
 >dbj|BAB73440.1| (AP003587) ORF_ID:all1741
 probable proteinase [Nostoc sp. PCC 7120]"
 /codon_start=1
 /transl_table=11

Query Match 9.3%; Score 369; DB 1; Length 10201;
 Best Local Similarity 53.3%; Pred. No. 3.3e-96;
 Matches 814; Conservative 0; Mismatches 695; Indels 18; Gaps 1;
 Qy 858 CATGACAAATACCAAAACCGTTATTGGTGTGATGGGATCAGGAGTGTCCGGCGCG 917
 Db 2372 CAAACAAATGGCAAGTTACTTTATCGCGGTAGACGTAGGACCGCGAGTGAAGAGCTGG 2313
 Qy 918 GATTTTGTATCATCAACGATCTGCTATCCCATGCCACAGAAATAATCAGACTACCG 977
 Db 2312 GGTATTCGATTGCGAGGCCGATGGTTGGCGAGCGCCGCGGTGAATATCATCATATCGC 2099

Qy 978 GCGCAGCGGAGCCGCGTGGGACACATCTCAGCCAGAGATCTGGCAGGCGGTCTGTTCATG 1037
 Db 2252 ACCCAAGGCTGATTTGTTGAGCAGTCTCCGAGAAATATTTGGCAAGCGGTGTGTAATGC 2193
 Qy 1038 TATTGTAATCGGCTCACTTGGCAGACGTTTGTGCCACAAAGTGTGGCAGCATCGGTTT 1097
 Db 2192 GGTGGGTGATCGGTTAATCAGCGCGATATCAACCCGATTCAGGTCAAAGGCGTGGT 2133
 Qy 1098 TGATGCCACCTGTTCTCTGGTGGTACTGGATAAAACCGTGTATCCATTCGCTGTTCAGCCC 1157
 Db 2132 TGATCGGACTTGTCTGCTGGTGGTGTGGATAAAGGAATCCACTCAGCGTTAGCC 2073
 Qy 1158 GGAAGGAGATCAAAAGCAAAATATCATTTGTGTGGATGATCACCGCGCCACCGAACAAG 1217
 Db 2072 TTCAGGCGCAATGAGCAAAATATGATTTGTTGGATGGACCCACCGGCGATCACTCAGCG 2013
 Qy 1218 AGAGGAATCAATGCCACTCACCCTCCGCTGCTGAACCTACGCTCGTGGTAAATTTCCGC 1277
 Db 2012 GGAGCGGATTAATGCCACCAAGCACCCTGCTGAGAGTTTGTGGTGGTGTGATTTCCGC 1953
 Qy 1278 TGAATGAAAACACCGAAAATTTCTCTGCTGAAAGAAAATATGCCAGAGATCTACGAAAG 1337
 Db 1952 AGAATGACAGACCAAAAGTTGTTGGTTAAAGCAACATATGCCGAACACGTTGAGTAA 1893
 Qy 1338 TCCCGGACAATTTTTCGATCTGGCGGATTTCTGACCTGGCGGGCTACCGGTGATTTAGC 1397
 Db 1892 TGTGCGCATCTGTTTATCAGATTTCTTACCTGGCGGTGCCACGAAAGATGAAC 1833
 Qy 1398 GGTTCAGTATGCATCTGTTTAAATGACGTGGCTGGCAGCATGAAATTCGCTGGGA 1457
 Db 1832 CCGTTCGCTCTGTTCCACCTCTGCAATGGACCTATCTGGGCCATGAAGATCGTGGGA 1773
 Qy 1458 TCCAGATTTATTCGCAACCATTTGGCTTTCGAGAGTTAGCGATGAAGATTTTATTCGCAT 1517
 Db 1772 TCCGTCTCTTCAAAATTTGGTGGGCTGGCGGATTTACTGGATAACACGCGGCGAAGAT 1713
 Qy 1518 TGGTCATCATATTTGTTTCTCCCGAAACACCTTTGTGGAATGGTTTTAACAGCAACAAGCCG 1577
 Db 1712 CGGTGCGAGCGTTAAACCATGCGGGCACCGTTAGGGCATGTTTTAAGTCAGCGGCGAC 1653
 Qy 1578 GCGAGAGATGGATTTACTTCCCGGACACCTGTCCCTGTAGGTTAATTTGATGCTCAGCG 1637
 Db 1652 ATCAGAAATGGGTTGATCCCGGTCACGCGGTGAGTGTGCGATTTATGATGCTCAGCG 1593
 Qy 1638 TGGTGGCATCGTACGGTTGGCGTGAAGAGTGGAGCGCTGAACAAT----- 1683
 Db 1592 TGGGACGATTGGCATCTCGGTGCCAGTGGTGTGAGGGGGGAAATGCTAACTTTGATCG 1533
 Qy 1684 ----CTCGCGTATGTTTTCGGCACTTCTTCATGACCATGGCATCTACACATTCCTCCCTC 1739
 Db 1532 GCGGATTGCGGTGATTTGGTGGCACCTCTACCGCCCATATGGCGATGTGCGCTCAGCACA 1473
 Qy 1740 GTTTGTACCGGGTGTCTCGGGGCGGTATACAGTGGGATGGTCCAGGTCTCTGGTGTAGT 1799
 Db 1472 CTTTATCATGTTGTTGGGGGCCATATTTATTCGGCCATTTTACCGGAATATTTGTTGAA 1413
 Qy 1800 TGAAGCGGGCAAAAGTGTCTCGGGGAGCAGTATTGACCAAGTCTACTTGTATTTCCATCCGCG 1859
 Db 1412 CGAGGCGGCGAGTCGGCAACGGGGGCTTTGATTGACCATATTTCAATCATCATCTCTTG 1353
 Qy 1860 TGTGGAAGAAGCTCGCAAAATGGCACAAACGTTGTGAATCAGCCCCCTCCCGCTGCGCTGC 1919
 Db 1352 CTATCCGCACTGCTGGAACAGGCTAAAAATAAGGGCGAAACCATCTATGAGGCGCTGAA 1293
 Qy 1920 TGATCGAATCTCGGAAAACCGGCAACCATCAGATCTGTGCGCTTGGCGGCAAGGCT 1979
 Db 1292 TTATATTTTGGCTCAGATGGCGGTGAACCTGAAAATATCGCGTTCTCCCTGACCAATGAT 1233
 Qy 1980 ACACGTGTGCGGAAATTTCTCGGAAATCGGCCCCCTTCGAGATCTCTCATGCGCAGAGC 2039
 Db 1232 TCATATGTTGCGGTACTTCCAGGTAAACGCTCTCCCGTGGCAACCACTCATGCGGG 1173
 Qy 2040 GGTAAATTTGGCCTGGGTATGGAGCGAGATCTGGATAATTTTACTCGCTTGTATATCGC 2099

Db	1172	CATTATTACGGCTGAAACTCTCTACCGCAGAGATATGGCGCTCGGTATTGGC	1113
QY	2100	TGGATTATCGGAATTTGGTTATGTCTGCGCCAAATTCCTCGAGCGCTCAACACAGCGCAGGG	2159
Db	1112	AACCATTCAGCGCGCTGGCTTGGCCACACGGCATATTATCGAAACCATGAATCAGATCG	1053
QY	2160	AGTAGTGAGTAAATATATCGTTATTAGCGCGGTGCGCGGACGATCCACTGTGATGACACA	2219
Db	1052	CTATAACATATCGACATATGATGGCCAGCGCGGAGCACAAGAACCCCATATTTGTACA	993
QY	2220	AATTCGTGGCAGATACCTGCGGTATTCGGTCTATTACCGCATGCTGCGAACCGGTTTT	2279
Db	992	GGAGCATGGAATGACCGGGCTGCCCATGCTGTTGCCAGAGAGAGTGGCGGATGCT	933
QY	2280	ATTAGCTCGGGCATCTTCCTGGTCTGCTGCGCGGAATATTGCACTCTCTGTTGGCGAAGC	2339
Db	932	GTTGGGAGTGCCATGATGGTACCGCTGCTGCTGGGTGTTGAGTCATTCGCCGAAGC	873
QY	2340	GATGCAACAAATTCACCCATGTGATAA	2366
Db	872	CATGGCGGCATGAGCCGCGCATCGTAA	846
RESULT 10			
AJ414158			
LOCUS			
Yersinia pestis strain CO92 complete genome; segment 18/20.			
AJ414158 AL590842			
VERSION			
AJ414158.1 GI:15981524			
KEYWORDS			
Yersinia pestis.			
SOURCE			
Yersinia pestis			
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
Yersinia.			
REFERENCE			
1 (bases 1 to 235050)			
AUTHORS			
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,			
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,			
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,			
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,			
Fellwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,			
Karlyshev,A.V., Moutle,S., Oyston,P.C.F., Quail,M., Rutherford,K.,			
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrerford,B.G.			
Genome sequence of Yersinia pestis, the causative agent of plague.			
Nature 413 (6855), 523-527 (2001)			
21470413			
11586360			
REFERENCE			
2 (bases 1 to 235050)			
AUTHORS			
Parkhill,J.			
Direct Submission			
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia			
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,			
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk			
Notes:			
Details of Y. pestis sequencing at the Sanger Centre are available			
on the World Wide Web.			
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).			
FEATURES			
source			
1. 235050			
Location/Qualifiers			
/organism="Yersinia pestis"			
/strain="CO92"			
/db_xref="taxon:632"			
/note="biovax: Orientalis"			
48. .830			
/gene="yPO3604"			
48. .830			
/gene="yPO3604"			
/note="no significant database hits"			
/codon_start=1			
/transl_table=11			
/product="hypothetical protein"			
/protein_id="CAC93073.1"			
/db_xref="GI:15981525"			
gene			
CDS			

/db_xref="SPTREMBL:Q8ZB25"			
/translation="MSDTPQHIIIRKTDPRNRPEFNAREINKINHPAREVNMGL			
IESALTFLRTHGVLDQFAVYVYTLARTQKNGLAGFTGECGLAGVVGQMDHLWPEQ			
QARSILIEWFTRVSNQLRQHDFTRDDLRVLRABRALQLLYDKQLQVLLKVPRIEN			
LLYLMMNTAKKLESADAAKQOATAPLKMPMPVYLSVPEAEPVRTAAAPPAANIE			
VRSPRRRLNGRCGCGALVPVSCVYVWRRLRFTLPSSNRCSK"			
827. .1360			
/gene="yPO3605"			
827. .1360			
/gene="yPO3605"			
/note="Similar to Arabidopsis thaliana F15p11.1 protein			
TR:Q9SRU2 (EMBL:AF160760) (187 aa) fasta scores: E():			
0.024, 31.5% id in 143 aa"			
/codon_start=1			
/transl_table=11			
/product="conserved hypothetical protein"			
/protein_id="CAC93074.1"			
/db_xref="GI:15981526"			
/db_xref="SPTREMBL:Q8ZB24"			
/translation="MTVIATQESAAALLWLNRPDVATYGEQLSTLENISPLVLNTAD			
OSVAMARQWSPDPSQVAESQWARKLVEARIGLAGTDSYFQLOQLHALSKLLEQE			
RSRGLTISYLKTAIVYQMTELNREIPLLELRQLAVSADEHQASPVLKIQIDDRWN			
ALLSRYHHLTQQTNSAR"			
1382. 3784			
/gene="yPO3606"			
1382. 3784			
/gene="yPO3606"			
/note="Similar to Escherichia coli VgrG protein TR:052672			
(EMBL:AF044503) (713 aa) fasta scores: E(): 0, 49.9% id in			
579 aa, and to Escherichia coli VgrE protein TR:052660			
(EMBL:AF044499) (702 aa) fasta scores: E(): 0, 49.3% id in			
576 aa. Similar to YPO3613 (78.0% identity in 803 aa			
overlap) and to YPO0764 (77.2% identity in 615 aa			
overlap)"			
/codon_start=1			
/transl_table=11			
/product="putative Rhs accessory genetic element"			
/protein_id="CAC93075.1"			
/db_xref="GI:15981527"			
/db_xref="SPTREMBL:Q8ZB23"			
/translation="MFAHDKANNNAADKTGTVTPTHTVSPAAPTVSGAOMTEMAGG			
IGDAVGGTAGKRINQAAEVAKTALDAKAKVLDGVTPTNIVATGAPTVSGAOMTEMA			
GGIGAVGGTAGKRINQAAEVAKTALDAKAKVLDGVTPTNIVATGAPTVSGAOMTEMA			
ISRLVKVSGSLQFTLTASLPQTFAVDFTLSEMLSPFLNVLGASADPAVDFAAV			
LDEDTALFIMREGVLORSTGNVAFSEQDGTGHTQTRYSMVIRPALWRTSLRNARIF			
QOASVEEITTLKENGNDFAFGFRHPHPVREFCVQESDEFITQRLTAEEGIFYF			
FEFSAGKNTVVYADVDSLPKGSALPYNPVAAQAEISITTTTSAQVVRPAWVKLD			
YTFKNPNAAFSEOSGELONORDPYEHFDPGRKDAQHGODFTRYRLDALRNDANL			
CGGASNDFTLQPGQLPSLYNHPRGDLNHAQWLLGLOHSGKQMALEQASGDOGTVLN			
HFSFIPHTQWRPTPLAKPMDGQPIAMVVGPPGEEIYCDGYRGLQLFLWDRYQGSN			
DNSSCWIRVTPWAGQWGLAIPRIGQEVVVDVFLHGDPDQPIVTRGTYHANNIPPS			
LPASKTQMAFRSKTHKGEGYNEMREFDAGGEGLEFMHAKQMDTIVLNDKRKDTVDQD			
SEHIGQDSVTVRNQSNITQNDREVEVTRDQQTQEVGNDYQLVWKEKEFTKIRYT			
EVHEDETTLVTKSIHAKQGDISTPNAGITITHDGAIVLQGVIRLAADMIDLNP			
EE"			
1421. 1441			
/gene="yPO3606"			
/note="PS00154 E1-E2 ATPases phosphorylation site."			
3038. 3103			
/gene="yPO3606"			
/note="Predicted helix-turn-helix motif with score 1023			
(72.67 SD) at aa 553-574, sequence IRLQLMDRIGQSNDSNCSWIR"			
3790. 4248			
/gene="yPO3607"			
3790. 4248			
/gene="yPO3607"			
/note="Similar to Pseudomonas aeruginosa hypothetical			
protein PA0094 TR:AA03484 (EMBL:AE004448) (144 aa) fasta			
scores: E(): 8.9e-06, 24.8% id in 141 aa"			
/codon_start=1			
/transl_table=11			
/product="conserved hypothetical protein"			
/protein_id="CAC93076.1"			
misc_feature			
misc_feature			
gene			
CDS			

```

/db_xref="GI:15981528"
/db_xref="SPTREMBL:O8ZB22"
/translation="MPENINTYCIQEGTLECEPGETDRTVNLFMKGPKCAAFALNIA
RDVPEAEVSDYVQRIQIGLKNLKYKVKHQIDYGAQHIFGPEVEATYMAETRR
WQORQVAIQVDHMIIFTATNASFFNPQOQATQEQWHSFAPWAREAGHV"
4241..7105
/gene="YPO3608"
CDS
4241..7105
/gene="YPO3608"
/notes="Similar to Escherichia coli Core protein TR:O52666
(EMBL:AF044501) (1420 aa) fasta scores: E(): 8.3e-32,
27.7% id in 824 aa, and to Escherichia coli RhsD protein
precursor SW:RHS_D_ECOLI (P16919) (1426 aa) fasta scores:
E(): 1.4e-28, 27.3% id in 842 aa"
/codon_start=1
/transl_table=11
/product="putative exported protein"
/protein_id="CAC93077.1"
/db_xref="GI:15981529"
/db_xref="SPTREMBL:O8ZB21"
/translation="MFEARVDDKLVHSSALAGFIIGSIIGAAVIFAAAAVAASIVLT
GGATLVATGFIQGMVTTLGVAVGLIIRSVGEKIGSMCHDVQIITGSKNVNSKR
AAHVELSTVACKDSDAIORMAEGSSNIFINSKAARLEKDTCTDAVVDSSNVTGG
GRVYIDLEKREIDEMRDLSEKLFIVAGLAGIFGAQKAGCFGLCKSLKALGEMAG
AAAGYLEKGAGIAFGYVPDVISQKLLTGGEDDTDFILPFIPLHWSRIYRSEN
HHVGLAQGWSLWERSLRKEDDSIVYQNDREIIVPLIKRGERVSPFTEHWLART
ERTVATISSPFCTCFEAFSEAGVAKLASLEDLNGHALYFSYDDIGLQKIKISTGY
GVYQVEKGLRYSVACVKGCTPGTLVRYOYNEOHLVSVNREGQITRORGYHGLIN
KLADRGLECRYWADLIGGTPRTHSNATNLGEQWDFYDIDNQOTLTLDTNQTACW
GNAQHLIDYRDFDGKYAFDYNLMPVRVVLAGEITLVLYDALARIQITDPLK
RETHIDYHNSLRVRRVYDPQGWKGEYDRTGLKENAPDGGVTLHYPGASSLPE
RITNAVAGTHLWGERGOLTEHTDCSGKLTREYDIDGHLTVIDAENHSTHYSVNR
LGQPTGRYADGRKEOLRYNAOGLVEQFDPDQVRLHWRVNLRGQPVSTFDRLQREVR
VRYDCHQTELDNANGGOYHFWSWGGOLVVEQYDNLVRRYVYCESGMMALETTA
PYDDLLTVSQQVSFDYDAGGRTQRLTGMSTATRYDWDIMBRLLEAERVYAVGEQAGI
VHGVDLYLADKAGHLHLESGDLGAVTQMDPLHLHLAALTLPDQGTLSWLYRGHVSA
IRHGDTLISSEFSRNDLHREVSRVTOGILTOYRDYDAMGRRL"
4298..4366
/gene="YPO3608"
/notes="one of 2 probable transmembrane helices predicted
for YPO3608 by TMHMM2.0"
4385..4453
/gene="YPO3608"
/notes="one of 2 probable transmembrane helices predicted
for YPO3608 by TMHMM2.0"
7133..8503
/gene="YPO3609"
7133..8503
/gene="YPO3609"
/notes="Similar to Escherichia coli hypothetical Core
protein RhsD TR:AAA24544 (EMBL:L19084) (1426 aa) fasta
scores: E(): 4.6e-14, 28.3% id in 473 aa, and to
Escherichia coli RhsE protein SW:RHS_ECOLI (P24211) (682
aa) fasta scores: E(): 1.6e-14, 26.8% id in 522 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC93078.1"
/db_xref="GI:15981530"
/db_xref="SPTREMBL:O8ZB20"
/translation="MAADLLPQGDIDWRKFSFDTAGELSMATDFIRGEQYRYDAEGR
LTDNRHQVLSVAEDFAYDNADNLLNLKLPFTDPLDTPVANNRLTQWQHYREY
DANGNTRHAGRMQHFAYDDNLLRLRWGTPGLGHDHSYRDLALGRHRSVTIK
RGAKEKTRQTFDFTQGLRLLEQHTDGNATYIYDPNESYTPPLARDHQHETESQVY
FHTDINCTPLDVTGEGKHKWCKYHAWGVTRQNVSDPRQSVRPAOPLRYGVYS
DDEGLHYNTFRYIDPEIGRFSFODPIGLAGGINLYOYGNPLTWIDPWGAWFGVD
TGSFDFPVKGSQNLNVEITMQAGRQDTEAFKLAGISKADATGYTWHHLNDFDPS
GKTMQLVTTSAHEATFPHAGSVSQSEKHEFLPSQSYGSDAVALSHSKGWLKRIPK
ALRSGC"
8505..8990
/gene="YPO3610"
8505..8990
/gene="YPO3610"

```

```

/notes="Similar to Bacillus subtilis Yokk protein TR:O31996
(EMBL:Z99115) (192 aa) fasta scores: E(): 0.043, 26.6% id
in 154 aa, and to Bacteriophage SPBC2 hypothetical protein
Yokk TR:O64025 (EMBL:AF020713) (192 aa) fasta scores: E():
0.043, 26.6% id in 154 aa"
/codon_start=1
/transl_table=11
Query Match 9.3% Score 369 DB 1 Length 235050;
Best Local Similarity 53.3% Pred. No. 4.8e-96;
Matches 81; Conservative 0; Mismatches 695; Indels 18; Gaps 1;
Qy 858 CATGACAAATAACCAAAACCGTTATTGTGATGTGGGATCAGGCAGTGTCCGCGCGG 917
Db 41868 CAAACAAATGCGCAAGTTACTTTATCGGCGTAGACGTAGGACCGGCGTGTGAGAGCTGG 41927
Qy 918 GATTTTGTATCAACGGATCTCTGCTATCCATGCCACAGAAAAAATACAGCATACGGG 977
Db 41928 GGTATTGCAATTTGCAGGGCCGTATGTTGGGCAGGCCAGCGGTGAATACCATGTTAA 41987
Qy 978 GCGCAGCGGAAGCCGCGTGGAAACAGTCCAGCCAGGAGATCTGGCAGCGGCTCTGTTTCATG 1037
Db 41988 ACCCAAGCTGATTTTGTGAGCAGTCTTCCAGAGAAATATTTGGCAAGCGGTGTGTAATGC 42047
Qy 1038 TATTGCTAATCGCTCACTCTGGCAGACGTTTGTGTCACAAAGTGTGGCAGGCATCGGTTT 1097
Db 42048 GGTGCGTGATCGGTTAATCAGGCCGATATCAACCCGATTCAGGTCAAAGGCTCGGTTT 42107
Qy 1098 TGATGCCACCTGTTCTCTGTTGTTACTCGATGATAAAGCGTATCCATTGTCCTGTCAGCCC 1157
Db 42108 TGATGCGACTTGTCTGCTGCTGTTGTTGGATAAAGAGGGAATCCACTGACGGTTAGCCC 42167
Qy 1158 GGAAGGAGATCAAGAAATATCATTTGTGTGGATGATCACCGCGCACCAAGCAACAGC 1217
Db 42168 TTCAGGGCGCAATGAGCAAAATGTGATTTGTTGGATGACCAACCGGCGATCAGTCAGGC 42227
Qy 1218 AGAGCGAATCAATGCCACTCACCATCCGCTGCTGAACCTACGTCGGTGGTGTAAATTTCCGC 1277
Db 42228 GGAGCGGATTAATGACCAACAGCACCCGCTCTGGAGTTTGTGCGTGTGTTGATTTCGCC 42287
Qy 1278 TGAATGAAACACCGAAATTCCTGCTGCTGAAAGAAATATGCCAGAGATCTAGCAACG 1337
Db 42288 AGAATGAGACACACAAAGTTGTTGTGTTAAAGCAACATATGCCGAACACGTTGAGTAA 42347
Qy 1338 TCCCGGACAATTTTCGATCTGGCGGATTTCTGACCTTGGCGGCTACCGGTGATTAGTAC 1397
Db 42348 TGTGCGCCATCTGTTTGTATCTACAGATTTCTTACCTGGCTGCCCAAGAGATGAAC 42407
Qy 1398 GCGTTCACTATGCACTGTTTACCTGTAATGAGCAGTGGCTGGCAGATGAAATTCGCTGGGA 1457
Db 42408 CCGTTGCGCTCTGTTCCACCGCTGCAAAATGGACCTATCTGGGCCATGAAGATCGCTGGGA 42467
Qy 1458 TCCAGATTATTTCCGCACCATTTGGCTTGCAGAGTTAGCGGATGAAGATTTTATTCGGAT 1517
Db 42468 TCCGTCTACTTCAATTTGGTGGGCTGGCCGATTTACTGGATAACACCGCGCGGAAGAT 42527
Qy 1518 TGGTCATCATATTTTCTCCCGGAACACCTTGTGGAATGGTTTAAACAGCACACACCGCG 1577
Db 42528 CCGTCCGCGGTTAAACCCATGGGGCACCGCTTAGGGCATGGTTTAAAGTCACGCGGCGAC 42587
Qy 1578 GCGAGATATGGATTACTTCCCGGCGCACACCTGTGCTGCTAGGTTTAAATGATGCTCAGC 1637
Db 42588 ATCAAAATGGGTTGATGATCCCGGCTACGCGGCTCAGTGTGCGATTTATTTGATGCTCAGC 42647
Qy 1638 TGGTGGCATCGTACGGTTGGCGTAGAGGTGGAGCGCTGAACAAT----- 1683
Db 42648 TGGGACGATTGGCATTTCCGGTGGCCAGTGGTGTGACGGGGGAAATGCTAACTTTGATGCG 42707
Qy 1684 -----CTCCCGTATGTTTTCGGGCACCTTCTTCATGCAACCATACCATCTTCCCTC 1739
Db 42708 GCGGATTCGCTGATTGTTGGCGCACCTCTACCGCCCATATGGCGATGTCGCGCTCAGCACA 42767
Qy 1740 GTTTGTACCGGGTGTCTGGGGGCGGTATTACAGTCGCGATGGTTCACAGGTCTGTGTTAGT 1799

```

Db	42768	CTTTATCACTGGTATTTGGGGCCATATTATTCGGCCATTTTACCGGAATATTTGGTTGAA	42827
Qy	1800	TGAAGCGGGCAAGGCTGCGGGAGCAGCTATTGACCAAGCTACTTGAATTCATCCCGC	1859
Db	42828	CGAGGCGGGCACTCGCAACGGGGCTTTGATTGACCATATTTATCATCATCCTTTG	42887
Qy	1860	TGTTGAAGAGCTCGGAAATGGCACAAGCTGTGAATCAGCCCTCCCGCTGGCTTGC	1919
Db	42888	CTATCGGCACCTGCTGGAACAGCTAAAAATAAGGCGCAACCATATGAGGCGCTGAA	42947
Qy	1920	TGATCGAATCTCGAAAAACGGCCCAACCATCAGATGCTGCGCCCTGCGCAAGGGCT	1979
Db	42948	TTATATTTTGGCTGAGATGGCGGGTGAACCTGAAATATTCGCGTTCTGACGAATGATAT	43007
Qy	1980	ACAGCTGGTGGCGGAATTTCTCGGAATCGCGCCCTTCGCGAGATCTTCATCCAGAGC	2039
Db	43008	TCATATGTTGCGGTACTTCACGGTAAACGGCTCTCCCGTGGCAACCAATCTGACGG	43067
Qy	2040	GGAATTTGCGCTGGGTATGAGCGAGATCTGGATAATTTACTGCGCTTGTATATCC	2099
Db	43068	CATTATTACTGCGCTGAAACTCTCTACACCGCAGAGATATGGCGCTCGCTTATTTGGC	43127
Qy	2100	TGATTTATCGGAATTTGTTATGTTGCTGCGCCCAATTTCTGAGCTCAACAGCGCAGG	2159
Db	43128	AACCATTTACAGGCGCTGGCGTTTGGCACACGGCATATTTATCGAAACCATGAATGG	43187
Qy	2160	AGTAGTGAGTAAATATCTGTTATTAGCGCGGTGCGCGGAGCATCTCACTGGTACGACA	2219
Db	43188	CTATAACATCGACATGATGCGCAGCGCGGAGCGGCACAGACCCCATATTTGTACA	43247
Qy	2220	AATCTGGCAGATACCTGCGGTATTTCCGGTATTTACCAGCAATGCTGCGAACCGTTT	2279
Db	43248	GGAGCATGCGAATGCGACGGCTGCGCCATGCTGTTGCGCAGAGAGTGGCGCATGCT	43307
Qy	2280	ATTAGCTCGGCATCTTTGTTGCTGCTGCGCGGAATATTTGACCTTCTGTTGGCGAAGC	2339
Db	43308	GTTGGGAGTGGCATGATGGGTACCGTCTGCTGCTGGGTGTTGAGTCACTCCCGGAAGC	43367
Qy	2340	GATGCAACAATTCACCATTTGGATAA	2366
Db	43368	CATGGCGGCATCAGCCGATCGGTAA	43394
RESULT 11			
BC014947			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
1 (bases 1 to 1923)			
Direct Submission			
Submitted (01-OCT-2001) National Institutes of Health, Mammalian			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
NIH-MGC Project URL: http://mgc.nci.nih.gov			
Contact: MGC help desk			
Email: cgapbs-remail.nih.gov			
Tissue Procurement: ATCC/DC/DTP			
CDNA Library Preparation: Rubin Laboratory			
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Genome Sequence Centre,			
BC Cancer Agency, Vancouver, BC, Canada			
info@bcgsc.bc.ca			
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,			
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,			

Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 34 Row: 1 Column: 7
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10433590.

FEATURES	source
Location/Qualifiers	
1..1923	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="MGC:22958 IMAGE:4871664"	
/tissue_type="Skin, melanotic melanoma, high MDR."	
/clone_lib="NIH_MGC_49"	
/lab_host="DH10B-R"	
/note="Vector: pOTB7"	
434..1753	
/codon_start=1	
/product="Unknown (protein for MGC:22958)"	
/protein_id="AAH14947.1"	
/db_xref="GI:15928959"	
/translation="MWLDHRAVSQVNRINETKHSVLYQVGVMSVEMOAPKLLWKEN LREICWDKAGHFDLPDLFSKATGVTGRTSLVCKWTYSAEKGDSDSPFKMIGLED EVADNYKIGNQVLPPLPGASLGNLTPEAROLDGLPIFGVWGPFSYFAPVGPWLNEG VRGHLICEGQPVTSRLAVICGTSCHMGISGRDLPFVGVWGPFSYFAPVGPWLNEG QSVTKLIDHMVQGHAAPELQVATKARCSYIAYLNHSLDIIKKAQPGFTVLDHV WPDHNRSLPLADLTGKMTGLKLSODLDDLIILATVQAIALGRTFTIEMAEAG HSISPLFLCGGLSKNPLFVQMHADITGMVPLVSQEVSLVGAALGACASGDFASVQ EAMAKMSKVGKVVPRLODKKKYKQYKVLKVEHQEYLAIMDD"	
BASE COUNT	486 a 422 c 520 g 495 t
ORIGIN	

...	Query Match	7.6%	Score 302.4;	DB 9;	Length 1923;
...	Best Local Similarity	52.1%	Pred. No. le-76;		
...	Matches 827;	Conservative	0;	Mismatches 716;	Indels 45; Gaps 5;
Qy	879	TATTGGTGTGATGTTGGGATCAGGCAGTGTCCGCGCGCGGATTTTGTATCTCAACCGATC	938		
Db	133	TCTGGGTGTGGAGCTTGGAAACAGGCAGTGTCCGTCGAGCTTGGTGACACAGATGGGT	192		
Qy	939	TCTGCTATCCCATGTCACAGAAAAATACGACTACGCGCGCGAGCGGAGCGCGCTGGA	998		
Db	193	CTGTGTTGGTCTTTTGACAGACCCCAATTAAGAAATGGGAGCCCCAGCTTCAACCCACCATGA	252		
Qy	999	ACAGTCCAGCCAGGAGATCTGGCAGCGGTCTGTTCATGATTTCTGTAATCGCGTCACTCT	1058		
Db	253	GCAGTCTCTCCGAGGACATCTGGCTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	306		
Qy	1059	GCACAGCTTTGTGCACAAGTGTGGCAGGATCGGTTTGTATGTCACCTGTCTCTGTCTGT	1118		
Db	307	ACAAGGATTTGATTTAAACCAAAATTCGAGGAGCTTGGTTTGTATGTCACCTGTCTCTGT	366		
Qy	1119	GGTACTGGATAAAACCGGTGATCCATTGCTGTCAGCCGCGAGGAGATGCAAGCAAAA	1178		
Db	367	TGTTTGGATAGCAGTTTCAACCCATTACAGTCAACAGGAGGGGATTCCTCATCGAA	426		
Qy	1179	TATCATTTGTGATGATGATCACCAGCCACCGCAACAGCAGGAGGATCAATGCCACTCA	1238		
Db	427	CGTCATCATGTGGCTGGACCATCGAGCAGTCAAGTAAAGTAAACAGGATCAATGAGACCAA	486		
Qy	1239	CCATCGGCTGCTGAACCTAGCTCGGTGGTAAAAATTTCCCTGAAATGGAACACCGCAAAAT	1298		
Db	487	GCACAGTGTCTCTCCAGTACGTGCGGGGGGTGATGTCTGTGGAAATGCAGGCCCGCAACT	546		
Qy	1299	TCTTGGCTGAAAGAAAATATCCAGAGAT---CTACGAACTGCGCGGAGCAATTTTCGA	1355		
Db	547	TCTTGGCTGAAAGAGAACTTTCAGAGAGATTTGCTGGGATAAGCGGGACATTTCTTTGA	606		

Db	1681	TAAGAAATACCAAGTAGTTCTGAAGCTG	1708
RESULT 12			
AX427495			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
826; Conservative			
0; Mismatches			
717; Indels			
45; Gaps			
5;			
QY	879	TATTGGTTGATGTGGGATCAGGCAGTGTCCGCCCGGGATTTTTCATCTCAACGGATC	938
Db	227	TGTGGGTGTGGAGCTTGGAAACAGGCAGTGTCCGTGACGCTCTGGTGGACAGAGTGGGGT	286
QY	939	TCTGCTATCCCATGCCACAGAAATAATCACGACTACCGCGCGACGGAGAGCCGCGTGA	998
Db	287	CTGTGGCTTTTTCGACACAGCCATTAAGATTGGAGGCCCGCAGTTCAACACCACATGA	346
QY	999	ACAGTCCAGCCAGGAGATCTGGCAGCGCTCTGTTTCATGTTATGCTTAATCGCTCACTCT	1058
Db	347	GCAGTCTCCGAGACATCTGGGCTGCGTGTGTGTTCACAAAGAAAGT-----TGT	400
QY	1059	GGCAGAGCTTTTTCGCACAAAGTGTGGCAGGATCGGTTTTCATGTCACCTGTCTCTCGT	1118
Db	401	ACAAGGATTTGATTTAAACAAATTCGAGACTTGGGTTTCATGCCACGTTCTCTCGT	460
QY	1119	GGTACTGGATTAACAAACGGTGTATCCATTGCTGTTCAGCCCGGAGGAGATGCAAGCAAAA	1178
Db	461	TGTTTGGATTAAGCAGTTTCACCCATTACCAGTCAACACAGTAAGGGGATTTCCCATCGAAA	520
QY	1179	TATCATGTGTGGATGATCACCGCCGCCACCAACACAGACAGGGAATCAATGCCACTCA	1238
Db	521	CGTCATCATGTGGCTGGACCATTGAGCAGTCAGTCAAGTTAACAGGATCAATGAGACCAA	580
QY	1239	CCATCCGGTGTCTGAATACGTCTGGTGTAAATTTCCGCTGAAATGGAACACACGAAAT	1298
Db	581	GCAGGTGTCTCCAGTACGTTCGGGGGGTGTGCTCTGTGGAATGACGGCCCGGAAACT	640
QY	1299	TCCTGGCTGAAGAAATATGCCAGAT---CTACGACGGTCCCGGACAAATTTTCGA	1355
Db	641	TCCTGGCTGAAGAGAACTTGAGAGAGATTGCTGGGATTAAGCGGGACATTTCTTTGA	700
QY	1356	TCCTGGCGGATTTCTGACCTCGGGGGCTACCGGTGATTTAGCGGTTTCAGTATGCACTGT	1415
Db	701	TCCTCGGACTCTTTATCTGTGGAGGCAACAGGTGTACAGACACGGTCTCTGCTCCT	760
QY	1416	TACCTGTAAATGGACGTGGCTGGCACAATGAAATTCGCTGGGATCCAGATTTATTTCCGCAC	1475

Db	761	GGTGTCTAAGTGACATATTTCAGCA---GAGAAAGGCTGGGACGACAGTTCCTCGAAAT	817
QY	1476	CATTGGCCTTGCAGAGTACGCGATGAAGATTTTATTCGATTTGGTCATCATATTGTTTC	1535
Db	818	GATTTGGTTTGGAGACTTTTTCAGATATAATACGCAAAATAGGAACCAAGTGTCTACC	877
QY	1536	TCCCGGAACACCTTGTGGAAATGTTTAAACAGACACAGCGCGGCGAGAGATGGATTACT	1595
Db	878	TCCTGGAGCTTCTCTTGGAAATGGCTCACACAGAGGACAGACCTTGGCCCTCT	937
QY	1596	CCCCGGCACACCTGTCGCTGTAGGGTTAATGATGCTCAGCTGTGGGATCGGATACGGT	1655
Db	938	CCCTGGGATTTGGCGTGCAGCTTCACTCAATGATGCCATGCGAGGACTAGGAGTGAT	997
QY	1656	TGCGCTAGAGGTTGGAGCGG-----TGAAACATCT	1685
Db	998	TGGGGCAGATGTAGAGGGCAGCGGCTCATCTGTGAGGGCAGCCAGTACGCTCAGCGGT	1057
QY	1686	CGCGTATGTTTTCGGCACTTCTTCATGCACCATGGCATCTTACCACCTTCTCCCTCGTTTGT	1745
Db	1058	GGCTGTCTATCTGTGGAAGCTTCTTGTGCATATGGGATCAGCAAGACCGATTGTTGT	1117
QY	1746	ACCGGTGCTCTGGGGCGGCTATTACAGTGGGATGTTTCAGGTCTGTGGTTAGTTGAAGG	1805
Db	1118	ACCAGCGCTCTGGGGCGCTTATTCTCAGCCATGGTACCTGGGTTCTGGCTGAATGAAGG	1177
QY	1806	CGGGCAAGTGTCTGGGAGCAGCTATTGACACAGTACTTGAATTCCTCCATCGGCTGTGA	1865
Db	1178	TGCTCAGAGGTTACTTGGAAATTTAGACACATGTTACAGGCGCATGCTCTTTTCC	1237
QY	1866	AGAAGTCTCGGAAATGGCACACGTTGTAATCAGCCCTCCCGTCTGGGCTTCTGATCG	1925
Db	1238	AGAACTACAAGTAAAGGCCACAGCCAGATGCCAGTATATATGCAAT---ATTGAACAG	1294
QY	1926	AATCTCTGAANAACGGCGCAACCATCAGATGCTGTGCGCTCGCGGAAAGGGCTACAGT	1985
Db	1295	TCACCTGGATCTGATTAAAGAGGCTCAGCGCTGGGTTTCTGTTACTGTGATTACATGT	1354
QY	1986	GGTCCGGAATTTCTCGGAAATCGCGCCCTTTCGCGAGATCTTCATGCCAGAGCGGTAAT	2045
Db	1355	TTGCCAGATTTCCATGCGCAACGGTCTCCCTTAGCAGATCTCACACTAAGGGCATGTT	1414
QY	2046	TTGTGGCTGGGTATGAGGCGAGATCTGGATATTTACTGCGCTTGTATATCCTGGATT	2105
Db	1415	CACCGGATTTGAAACTGTCTCAGGACCTTGATGATCTTGCCATCTCTACCTGGCCACAGT	1474
QY	2106	ATCGGAATTTGTTATGGTCTGCGCAAAATCTCGAGCTCAACAGCGGAGGAGTAGT	2165
Db	1475	TCAGGCATTTCTTGGGACTCGCTTCATATAGAGCCATGGAGCAGCAGGCACTC	1534
QY	2166	GAGTAAATATCTGTTATTAGCGCGGTGCGGGCGAGCATCTGTTAGCAGCAAAATCT	2225
Db	1535	AATCAGTACTCTTTTCTATGTGGGCGCTCAGCAAGAATCCCTTTTGTGCAATGCA	1594
QY	2226	GGCAGATACCTCGGTATTTCCGTTCATTACCAGCAATGCTGCAACCGGTTTATTTAGG	2285
Db	1595	TGCGGACATTTACTGGATGCTGTGTGCTGTGTCGCAAGAGGTGGAGTCCCTTCTTGGG	1654
QY	2286	CTCGGCATTTCTGTTGCTGCTGCGCGAAATATTGCACCTTCTGTTGGCGAAGCATGCA	2345
Db	1655	TGCTGTGTTCTGGTGGCTGCTGCTCAGGGGATTTGCGTCTTCTACAGGAACATGGC	1714
QY	2346	ACAATTCACCATGTGGATAAATATTATTATCCGCAAGACGCTATCAATCTCTTCTATCA	2405
Db	1715	AAAAATGAGCAAGTTGGGAAAGTTGTGTTCCTCCGAGACTACAGGATAAAAAATACTATGA	1774
QY	2406	TCGTCCATATGAGGCTTATAGCAGTTG	2433
Db	1775	TAAGAAATACCAAGTATTCTCTGAAGCTG	1802

RESULT 13

AK090568	Homo sapiens	cdna FLJ33249 fis	2485 bp	mrna	linear	PRI 15-JUL-2002
LOCUS	to L-RIBULOKINASE (EC 2.7.1.16).					
DEFINITION	AK090568					
ACCESSION	AK090568.1	GI:21748754				
VERSION						
KEYWORDS	oligo capping; fis (full insert sequence).					
SOURCE	Homo sapiens normal astrocytes (NHA5732) cdna to mRNA,					
	clone lib:ASTRO2 clone:ASTRO2005081.					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.					
TITLE	NEDO human cDNA sequencing project					
JOURNAL	unpublished					
REFERENCE	2 (bases 1 to 2485)					
AUTHORS	Isogai,T. and Yamamoto,J.					
JOURNAL	Direct Submission					
COMMENT	Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.					
FEATURES	Location/Qualifiers					
source	1. .2485					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="ASTRO2005081"					
	/cell_type="normal astrocytes (NHA5732)"					
	/clone_lib="ASTRO2"					
	/note="cloning vector: pME18SFL3-primary culture, normal astrocytes"					
BASE COUNT	618 a 550 c 605 g 712 t					
ORIGIN						
Query Match	7.1%;	Score 282.6;	DB 9;	Length 2485;		
Best Local Similarity	52.3%;	Pred. No. 6.8e-71;				
Matches	776;	Conservative 0;	Mismatches 664;	Indels 45;	Gaps 5;	
QY	879	TATTGGTGTGATGTGGGATCAGCAGTGTCCGCGCGGGATTTTGTGATCTCAACGGATC	938			
Db	206	TGTGGTGTGGAGCTTGGACAGCAGTGTCCGTGCGAGCTCTGGTGACAGGGTGGGGT	265			
QY	939	TCGTCTATCCATGCCACAGAAAAATCACGACTACGCGCGCGAGCGGAGCGCGTGA	998			
Db	266	CCTGTTGGCTTTTGCAGACCAGCAATTAAGAAGTGGAGCCCGCAGTTCAACACCACATGA	325			
QY	999	ACAGTCCAGCCAGAGATCTGGCAGCGGTCTGTTCTATGTATTGCTAATCGGCTCACTCT	1058			
Db	326	GCAGTCTCCGAGACATCTGGGCTGCGTCTGTGTTGTCACAAAGAAAGT-----TGT	379			
QY	1059	GGCAGAGCTTTGTGCACAAAGTGTGGCAGGCATCGGTTTGTGATGCCACCTGTTCTCTGGT	1118			
Db	380	ACAAGGATGATTTAAACCAAAATTCAGGACTTGGTTTGTATGCCACGCTGTTCTCTGGT	439			
QY	1119	GGTACTGGATAAAACGGTGTATCCATTGCTGTACGCCCGGAGAGATCCAAAGCAAAA	1178			
Db	440	TGTTTTGGATAAGCAGTTTCAACCATTTACCACTCAACACAGGAAGGATTCCTCCATCGAAA	499			
QY	1179	TATCATTTGTGTGATGATCACCGCGCCAGCAACAGCAGAGCAATCAATGCCACTCA	1238			

[illegible]

5.78; Score 226; DB 10; Length 1842;


```

1585..2304
/gene="Atu4452"
CDS
1585..2304
/gene="Atu4452"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45246.1"
/db_xref="GI:17742930"
/translation="MRGTGGLLFDGCVLVDSPEVSIIVLLDMLSHLGVMTGEEEA
YERLGSVASMTTLFEYEGVETDIDFLDMRATLFRFRTKLPIDGIAETDELL
PLKRVASSOPERRIRYSLGLIDKFEPHFVFSATVVRNGKPADPLFLHAARSMGVE
PRQCIVEDSPAGIAAKAGMGVFAFTGGSHARFPAPREXIAIGLAGADAVDPMDLV
QLVGSYVGRGGFDQVERDAG"
2359..3936
/gene="Atu4453"
2359..3936
/gene="Atu4453"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ribitol kinase"
/protein_id="AAL45247.1"
/db_xref="GI:17742931"
/translation="MRONLVAVDVGTSARAGIDPAGRIILARSIHPILMORPRENH
AEHSTDIVNACVIAKALADAGVLQSIAGIFDATCSLVIRDERGEPVSTDD
DRFDITVLDHRAIGEADRLTASGHRVLDVFNAGSVPEMQPKMLWKLTHMPQSWSRM
SFADLADFLTKATGSNORNCQTAKNFWLAQEPHGWQADYLAFAGLDDLKERAGL
PETVMFGAGTGLPSPEAAELGLDTGCVQAAGMIDAYAGALGAGCLADEVDGKHA
LIAGTSCLVAMTQPMPSLWGPYWOAVLPHWLVGGOSATGALLDHIIVRMAAG
GEPTALHARIVARVTELRELEGAFAFADRLHVPDFHGNRSPADPHAVGVVSGLTLD
TSFSLCLRYRTAIALGARHVLDMAREGYAVETLHVTTGGVHKNPMLLMELADVT
KRIIVPATADAVILGTAMTAAATAGGVHASLAAGAAMPNGAEISGNPALAAHYERD
YRRFLAMIRHQELES"
complement(4041..4874)
/gene="Atu4454"
complement(4041..4874)
/gene="Atu4454"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, RpiR family"
/protein_id="AAL45248.1"
/db_xref="GI:17742932"
/translation="MVEKLLNLTRANKAGTPSPERKIAKYLIEHLDELFFETAOTLAA
KLUSPMTVGRFLSLGYRQFSDIRADLRHAETVVGADQPVTAEGGRHSNPFQQLS
QIQIAIQTFMTQTQIWRLANNETTSADDVCLATSPGLDAGRHFYRLLECRGHIH
YLGSDNAYVALWDYDIPARTLLVIMDCGGLLPQRLSTTARKGYRTLITTFYEW
GPESVDLCLAMPQSONGGQLQLVSLLEFTLCLASAGADDAGRKARVKNLAALKRSLR
S"
4897..5067
/gene="Atu4455"
4897..5067
/gene="Atu4455"
/note="putative; ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL45249.1"
/db_xref="GI:17742933"
/translation="MNGDAFARPCRKANHDLTVETSPFCHVLRHRTGMRIIVDGLLYFR
AIFYSRIPEFV"
5170..5907
/gene="Atu4456"
5170..5907
/gene="Atu4456"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45250.1"
/db_xref="GI:17742934"
/translation="MIFDAARLAECLNFNAETRSVFWKVLGTLVLVLAALFAIRSF
IYFALPWIDTLIPGVPDAGWLSFVFAIFAGIGLALALALLISPTAVTAGLFLDDVA
DVVEKQNPDDPPGKAMPIDGEAVLSSRIAGNLVALLLLVPGVNLITFAFFLVN
GYLLGREFFFAAMRFRPADARLFRVHRKHTVFNAGLVIAFLAVPFLNLITPLFAA
SLMVHLHKAVSRROPSFAAGGTEQLRG"
complement(5869..6579)
/gene="Atu4457"
complement(5869..6579)
/gene="Atu4457"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="endonuclease III"
/protein_id="AAL45251.1"
/db_xref="GI:17742935"
/translation="MRVKYASKDELTEIFRRFSIORPEPKGELEHTNPFITLVAVAL
SAQATDGVNRATRALFKVADTPKMLALGEOLIGHIKTIGLYRNKAKNVIALSQML
IDNFGGEVPKTRRELYTLPGVGRKTANVVMNAFGVPTLAVDTHVFRANRLCLAPGK
TTDEVEDRLRIIPQYLFAHHLWILHGRYCCCKARKPECRCVADIACKSPEKTFDI
PAPLVELPAQLFGAAGG"
6689..7171
/gene="Atu4458"
6689..7171
/gene="Atu4458"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45252.1"
/db_xref="GI:17742936"
/translation="MDTLNDQPIFAELVPHRSRGRFRLLIITGTACLIYGGFFL
ATGAMPILGFLGDLFLYIAFRANRAAKAREVSVSTNLSIRKFSFAGRMVYRF
NPFWARFRVRHDEIGIVSMHVTGGRATDIGSFLNPDDRSFAKAFGALATVRRRM"
7438..8310
/gene="ada"
7438..8310
/gene="ada"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="6-O-methylguanine-DNA methyltransferase"
/protein_id="AAL45253.1"
/db_xref="GI:17742937"
/translation="MNANTMLNEDITPIGSDYDVTVRGVIELLTDYREOPSLAATAAR
LGQSPQLOKQFTFRWAGLSKRAFLQAVILDHAKRLRLREDDLPLETSEVMSGFGRLL
HDLFVTHEAMSPGEMKAGGGLTIRYGFHASPFGLALVMTDRGLAGCAFDGDERA
CFEDMAGRPNADYVEDREATAFYAARIFEPAMWTADPLRVLLGTDFQVRWKSLL
KIPMGVATYVNIACDIIQGTASRAVGAAGVNPVSFVVPCHRAVKGSLGTGYHML
TKRAMLGWETKA"
8375..8599
/gene="ada"
8375..8599

```

```

Query Match      5.5%; Score 218.6; DB 1; Length 11164;
Best Local Similarity 49.3%; Pred. No. 5.2e-52;
Matches 714; Conservative 0; Mismatches 694; Indels 39; Gaps 4;

QY 872 AAACCGTTATTGCTTGTGATCGGATCAGGAGTCCGCCGCCGGGATTTTGTGACTCA 931
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2369 AAAATCTCGTCGCTGTCGATGTCGGAGCGCGCGCGGATTTTCGATCCAG 2428
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 932 ACGGATCTCTGCTATCCCATGCCACAGAAAAATACGACTACGCCGCGCAGCGGAAGCC 991
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2429 CCGGCCGGTGTGCGCGCGCTCCATTCATCCCATCTCATGACGCGCGCGCGGAAACC 2488
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 992 GCGTGGAACAGTCCAGCCAGGAGATCTGCAGCGGCTGTGTTTCATGTATTCGTAATGCGC 1051

```

```
Db 2489 ACCCGAGCAGATCCACCGATATATGGAATGCCGTCTGCATTCGCGTGAAGCGCGCG 2548
QY 1052 TCACCTGGCAGACGTTTGTGCAAAAGTGTGGCAGCATCGGTTTGTATGGCCACCTGTT 1111
Db 2549 TTCCCGATCGCGGGTCTGCTCCGCAAGCATTCGCCCATTTGAGCCACCTGCT 2608
QY 1112 CTCTGGTGTACTGGATATAAAACGGTGTATCCATTGGCTGTACGCCGGAAGAGATGCAA 1171
Db 2609 CGCTGGTTCATCCCGATAGCGGGGGAGCCGTTTCGGTTTCACGACCGATGACGACC 2668
QY 1172 AGCAAAATCATCTGTCTGATGATCACCGCGCACCCGAACAGCAGACGAGATCAATG 1231
Db 2669 GTTTCAGACCATCTGCTGTGCTTGACACCGCGCCATCGCGAGCGCGACCGGCTGACGG 2728
QY 1232 CCACCTCACCATCGGTGCTGAATAGTGTGGTGAATTTTCGCTGAAATGGAACAC 1291
Db 2729 CCTCGGGCCATCGGTGCTCGATTTCCCGGCAACAGCGTCTCGCCGGAATGCAGATGC 2788
QY 1292 CGAAATTCCTCGGCTGAAGAAATATGCCAGAGATCTACGAACGTCGCGGACATTTT 1351
Db 2789 CGAAGCTGATGTGGCTGAAGACGCATATGCCGCAAGCTGGTCGCGCATGTCTTCGCT 2848
QY 1352 TCGATCTGGCCGATTTCTGACCTGCGGCTACCGTGATTTAGCGCTTCAGTATGCA 1411
Db 2849 TCGATCTCGCGGATTTCTGACCTGGAAGCGACCGGCTCGGCACAGCGCTCCAATTGCA 2908
QY 1412 CTGTTACCTGTAATGGACGTGCTGGCAGATGAAATGCTGGGATCCAGATTAATTTCC 1471
Db 2909 CGCAGACGGGAAATGGAATCTCTGCGCAGGAACATCCCGCTGGCAGCGGATATC 2968
QY 1472 GCACCATTTGCGCTTGAGAGTTAGCGGATGAAGATTTTATTCGCATTTGTCATCATTTG 1531
Db 2969 TGGCTTTTGGCGACTGGAGACCTGAAGAGCGGCGAGGCTGCTGAAACTACCGTCA 3028
QY 1532 TTTCTCCCGAACACCTTCTGGAATGGTTTAACACACAGCCCGCGGAGAGATGGAT 1591
Db 3029 TGGCGGGCGGCGCATCGG-----CCCGCTTTGCGCCGAGCGCCGCGCAGAACTTGGTC 3082
QY 1592 TACTCCCGGCGACACCTGTGCTGTAGGGTTAATGTATGCTACGCTGTGGTGGCATCGGA 1651
Db 3083 TCGATACCGGCTCCAGGTGGCGCGGCATGATCGATGCTATGCCGCTGCGCTCGGG 3142
QY 1652 CGGTTG-----CGGTAGAGGTGGAGCGCTGAACAACTCGCGGTATGTTTTCGGCACAT 1705
Db 3143 CTCTGGGTGGATGCTTGGCGAGGATGTCGGTAAACATGTGCGGCTCATCGCGGCACT 3202
QY 1706 CTTTCATGCACCATGGCATCTACCACTTCTCCCTGTTTGTACCGGCTGTCTGGGGCCGT 1765
Db 3203 CGAGTTGCTTGTGCGCATGTCCACGACCGGATGCCGCGGCGAGCCTGTGGGGACCCT 3262
QY 1766 ATTACAGTGGATGGTTCCAGGTCTGTGGTTAGTTGAAGCGGGCAAAAGTGTCTGCGGGAG 1825
Db 3263 ACTGGCAGGGGGTGTGCTTGGCCCTGGCCACTGGCTGTGGAGGGTGGCAATCGGCTACCGGG 3322
QY 1826 CAGCTATTGACAGCTACTTGTATTTCCATCCGCGCTGTGAAGAAAGCTCGCGAAATGGCAC 1885
Db 3323 CGCTGTCTGACCATATCGTGGCATGATCGCGGGTGGCGA-----GCCGG 3370
QY 1886 AACGTGTGAATCAGCCCCCTCCCGCTTGGCTTGTCTGATCGAATCTCGAAAAACGGCG 1945
Db 3371 ATACGGGTTGATGCGCCGATGCTTGGCGCGCTCACTGAATTTGCGCGAATCGGAGGGG- 3429
QY 1946 AACCATCAGATGTGTGGCCCTGGCGAAAGGGCTACACGTGGTGGCGGAATTTCTCGGA 2005
Db 3430 -----GAGGCTTTGGCGGACCGGCTGATGTTCTGCTGACTTCCACGGCA 3475
QY 2006 ATCGCGCCCCCTTCGCAGATCTCATGCCAGACGGGTAATTTGTGGCCCTGGGTATGGAGC 2065
Db 3476 ACCGTTTCGCGCTTGGCGATCCGATCGCGTCTGGCGTCTGAGCGGCTGACGCTCGATA 3535
QY 2066 GAGATCTGGATATTTACTCGGCTTCTATATCGCTGGATATCGCGGAATTTGTTATGTC 2125
```

```
Db 3536 CTTCTCTCGACAGTCTCTGCTGCCGTCTTTATTTGGCGCACCGCGATCGCGATCGCGCTCGGCG 3595
QY 2126 TGGCCCAAATTTCTGAGCGCTCAAAACAGCGCAGGAGTAGTAGTAAATAATATCGTTATTA 2185
Db 3596 CGCGCCACGCTGCTCGATGCCATGGAGCGTTTCGGTTATGCGGTGGAACCGCTGCATGTTA 3655
QY 2186 GCGGCGGTGCGGCGAGCATCCACTGTGTACGACAAATTTCTGGCAGATACCTGCGGTATTC 2245
Db 3656 CCGCGGGCATGTGAAGAACCGCTGTTGATGGAGCTTTACGCCGATGTGACGCGCAAGC 3715
QY 2246 CGGTCAATTACACGCAATGCTGCGAAACCGGTTTTATTAGGCTCGGCCATTTCTTGGTCTG 2305
Db 3716 GCATCGTGGTTCCGCGACCGCGATGCGGTGCTGCTCGGCACGGCGATGACGCGCGCA 3775
QY 2306 TCGCCGG 2312
Db 3776 CGGCCGG 3782
```

Search completed: March 21, 2003, 20:42:22
Job time : 11847.3 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:12:47 ; Search time 881.004 Seconds
(without alignments)
10150.564 Million cell updates/sec

Title: US-09-802-208B-2
Perfect score: 3971
Sequence: 1 atcgattgagcagttgctt.....tcacgcgcgcagccatcgat 3971

Scoring table: IDENTITY NUC

Gapop 10.0., Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3971	100.0	3971	22	AAD16811
2	311	7.8	1401	24	ABN90720
3	302.4	7.6	1876	24	AAD38854
4	302.4	7.6	2050	22	AAI59449
5	300.8	7.6	1991	22	AAS03016
6	211.6	5.3	2602	23	ABL26431
7	183	4.5	1947	22	AAH15580
8	177.4	4.5	5019	23	ABL26430
9	124	3.1	1044	22	AAH14791
					Escherichia coli s
					Staphylococcus epi
					Human kinase (PKIN
					Human polynucleoti
					Human diagnostic a
					Drosophila melanog
					Human cDNA sequenc
					Drosophila melanog
					Human cDNA sequenc

10	112.2	2.8	571	22	AAI61235
11	104.2	2.6	795	22	AAH04701
12	90.4	2.3	677	24	ABN60111
13	86.2	2.2	409	21	AAC02752
14	82	2.1	613	22	AAH07400
15	78.4	2.0	640	21	AAF11917
c	72.2	1.8	1788	18	AAV74641
16	71.6	1.8	420	22	AAK87911
17	71.2	1.8	1335	22	AAH65092
c	71.2	1.8	34980	22	AAH64966
18	70.8	1.8	777	24	ABN70575
19	69.8	1.8	601	21	AAF08038
20	69.6	1.8	696	24	ABN92097
c	69	1.7	414	22	AAI08830
21	68.8	1.7	744	22	AAI14510
c	68.8	1.7	1020	16	AAO67531
22	68.8	1.7	11604	22	AAI14501
c	68.8	1.7	15079	16	AAQ91580
23	68.8	1.7	15079	22	AAI14499
c	68.6	1.7	394	23	ABV36510
24	65.8	1.7	413	22	AAI19430
25	64.8	1.6	792	24	ABN66664
26	64.6	1.6	789	23	AA52863
27	64.6	1.6	29729	20	AAI13175
28	63.4	1.6	468	24	ABK74996
29	63.2	1.6	744	24	ABK79267
30	61.2	1.5	2687	21	AAZ30700
31	60.8	1.5	732	24	ABN68642
32	60.8	1.5	2155561	24	ABN71527
33	59.2	1.5	732	24	ABN67976
34	58	1.5	807	22	AAH52045
35	57.6	1.5	728	21	AAC77209
c	57.2	1.4	2944528	24	ABA03041
c	56.6	1.4	469	23	ABV06559
36	56.4	1.4	8320	22	AAS02322
37	56.4	1.4	8320	24	ABK54204

ALIGNMENTS

RESULT 1
AAD16811 standard; DNA; 3971 BP.

AC AAD16811;

XX 29-NOV-2001 (first entry)

DE Escherichia coli strain C ribitol operon (rtl operon).

XX Positive selection system; metabolise; arabinol; ribitol; mannitol;

KW transgenic cell; marker gene; ribitol dehydrogenase; ribitol kinase;

KW ribitol transporter; rtl operon; ds.

XX Escherichia coli C.

XX Key Location/Qualifiers

FT CDS 96..848

FT CDS /tag= a

FT CDS /product= "E. coli ribitol dehydrogenase"

FT CDS 859..2463

FT CDS /tag= b

FT CDS /product= "E. coli ribitol kinase"

FT CDS 2565..3839

FT CDS /tag= c

FT CDS /product= "E. coli ribitol transporter"

WO200166779-A2.

13-SEP-2001.

08-MAR-2001; 2001WO-US07474.

XX 08-MAR-2000; 2000US-0189291.
PR 15-AUG-2000; 2000US-0255595.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX Parrott W, Lafayette P, Kane P;
PI
XX WPI; 2001-565596/63.
DR P-PSDB; AAE09779, AAE09780, AAE09781.
XX
XX Positively selecting transformed cells comprising selectable marker
PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinotol, ribitol which confer selective advantage on transformed
PT cells -
XX
XX Claim 1; Page 35-36; 37pp; English.
XX
CC The present invention relates to a positive selection system that
CC involves conferring to transferred cells the ability to metabolise
CC arabinotol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C ribitol operon (rbl operon). This sequence encodes ribitol
CC dehydrogenase, ribitol kinase and ribitol transporter.
XX
SQ Sequence 3971 BP; 919 A; 930 C; 1020 G; 1102 T; 0 other;

Query Match 100.0%; Score 3971; DB 22; Length 3971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGATTGAGCAGTTGCTTACACGGCAGCTAAATCCCGTTTCAGTGGCGTCAAGCAAA 60
DB 1 ATCGATTGAGCAGTTGCTTACACGGCAGCTAAATCCCGTTTCAGTGGCGTCAAGCAAA 60
QY 61 CAGAGACTATAAATTCGCCCTCGTAAAGGATTATATGATGATGATGATGATGATGATGAT 120
DB 61 CAGAGACTATAAATTCGCCCTCGTAAAGGATTATATGATGATGATGATGATGATGATGAT 120
QY 121 TGAATCTCCCTTAATGCAAGTTGCGAGCTATCACTGGCGCTGGTCAAGTATGGCC 180
DB 121 TGAATCTCCCTTAATGCAAGTTGCGAGCTATCACTGGCGCTGGTCAAGTATGGCC 180
QY 181 TCGAATGTGCAAAACGCTGCTCGATGAGGAGCAAGTAGTATTGATTGACCGGGAAG 240
DB 181 TCGAATGTGCAAAACGCTGCTCGATGAGGAGCAAGTAGTATTGATTGACCGGGAAG 240
QY 241 GCGCAAACTGCACAAGATTGTCGTGAGTTAGGCGGAAACCGCTACGGCTGCAACTCG 300
DB 241 GCGCAAACTGCACAAGATTGTCGTGAGTTAGGCGGAAACCGCTACGGCTGCAACTCG 300
QY 301 ATCTCTTCAATTAATCAGCAAGTCGATACATGCTGGCGACATTATCGAACTGGCGGGTG 360
DB 301 ATCTCTTCAATTAATCAGCAAGTCGATACATGCTGGCGACATTATCGAACTGGCGGGTG 360
QY 361 GGCTGGATATTTTTCATGCCAATCAGGCGCTTATATTGGCGGCGCAGTGGCTGAAGGTG 420
DB 361 GGCTGGATATTTTTCATGCCAATCAGGCGCTTATATTGGCGGCGCAGTGGCTGAAGGTG 420
QY 421 ATCCAGATGCTGGGATCGTGTGTTAAATCTGAATATAAATCCGGCGTTTCGCTGTGCC 480
DB 421 ATCCAGATGCTGGGATCGTGTGTTAAATCTGAATATAAATCCGGCGTTTCGCTGTGCC 480
QY 481 GTGCAGTCTGCCGATATGATTGGCAGAGGTTCGGCGCATATAATTTTACCAGTTCCA 540

DB 481 GTGCAGTCTGCCGATATGATTGCGCAGAGGTTCGGCGCATATAAATTTTACCAGTTCCA 540
QY 541 TCGCGGGCGTTCGGTTTATCTGGGAACCGATCTACACCGCGTCCAAATTTGCCGTTTC 600
DB 541 TCGCGGGCGTTCGGTTTATCTGGGAACCGATCTACACCGCGTCCAAATTTGCCGTTTC 600
QY 601 AGGCATTCGTACACACTACCCGCCCGCAGGTTTCTCAATATGCGGTGCGGTGCGG 660
DB 601 AGGCATTCGTACACACTACCCGCCCGCAGGTTTCTCAATATGCGGTGCGGTGCGG 660
QY 661 TGCTGCCAGGACAGTAGTACCTGCTGCTGATGACTGGCCAAAGCAAAATGGAAG 720
DB 661 TGCTGCCAGGACAGTAGTACCTGCTGCTGATGACTGGCCAAAGCAAAATGGAAG 720
QY 721 AAGCCCTGCGCAAAATGCTGATGCAACCGATTCAAGTGGCGGAATCAGTATCTTTA 780
DB 721 AAGCCCTGCGCAAAATGCTGATGCAACCGATTCAAGTGGCGGAATCAGTATCTTTA 780
QY 781 TGGTGACCCGCTCGAAAAATGTCACCGTGCAGATTTAGTATCTGCTGCGCTGGCAGTCTG 840
DB 781 TGGTGACCCGCTCGAAAAATGTCACCGTGCAGATTTAGTATCTGCTGCGCTGGCAGTCTG 840
QY 841 ATCTGTAAGGGCGCAATCATGACATAAACCAAAACCGTTATTGGTGTGATGTTGGATCA 900
DB 841 ATCTGTAAGGGCGCAATCATGACATAAACCAAAACCGTTATTGGTGTGATGTTGGATCA 900
QY 901 GGCAGTGTCCGCGCGGGGATTTTGTATCTCAACGGATCTCTGCTATCCCATGCCACAGAA 960
DB 901 GGCAGTGTCCGCGCGGGGATTTTGTATCTCAACGGATCTCTGCTATCCCATGCCACAGAA 960
QY 961 AAAATCAGACTACCGCGCGCAGCGGAGCGCGTGGAAACAGTCCAGCCAGGAGATCTGG 1020
DB 961 AAAATCAGACTACCGCGCGCAGCGGAGCGCGTGGAAACAGTCCAGCCAGGAGATCTGG 1020
QY 1021 CAGGCGGCTGTTTCATGATTGTAATGCGCTCACTCTGGCAGACGTTTGTGCACAAAGT 1080
DB 1021 CAGGCGGCTGTTTCATGATTGTAATGCGCTCACTCTGGCAGACGTTTGTGCACAAAGT 1080
QY 1081 GTGGCAGGATCGGTTTGTATGCCACCTGTTCTCTGGTGTACTGTGATAAAACGGTGAT 1140
DB 1081 GTGGCAGGATCGGTTTGTATGCCACCTGTTCTCTGGTGTACTGTGATAAAACGGTGAT 1140
QY 1141 CCATTGCTGTGTCAGCCCGGAGAGATGCAAGCAAAATATCATTTGTGTGATGATGATCAC 1200
DB 1141 CCATTGCTGTGTCAGCCCGGAGAGATGCAAGCAAAATATCATTTGTGTGATGATGATCAC 1200
QY 1201 CCGCGCCACCGCAACAGCAGCAGCAATCAATGCCACTCACCATCCCGTGTGAACCTAGTC 1260
DB 1201 CCGCGCCACCGCAACAGCAGCAGCAATCAATGCCACTCACCATCCCGTGTGAACCTAGTC 1260
QY 1261 GGTGTGTAATTTTTCGCTGAAATGGAACACCGCAAAATTTCTCTGGCTGAAAGAAAATATG 1320
DB 1261 GGTGTGTAATTTTTCGCTGAAATGGAACACCGCAAAATTTCTCTGGCTGAAAGAAAATATG 1320
QY 1321 CCAGAGATCTACGACAGTCCCGGACAAATTTTTCGATCTGGCGGATTTTCTGACCTGGCGG 1380
DB 1321 CCAGAGATCTACGACAGTCCCGGACAAATTTTTCGATCTGGCGGATTTTCTGACCTGGCGG 1380
QY 1381 GCTACCGGTGATTAGCGGTTTCAGTATGCATGTTTACCTGTAAATGGAGCTGGCTGGCA 1440
DB 1381 GCTACCGGTGATTAGCGGTTTCAGTATGCATGTTTACCTGTAAATGGAGCTGGCTGGCA 1440
QY 1441 CATGAAATTCGCTGGGATCCAGATTATTTTCCGACCATTTGGCCTTCAGAGTTTAGCGGAT 1500
DB 1441 CATGAAATTCGCTGGGATCCAGATTATTTTCCGACCATTTGGCCTTCAGAGTTTAGCGGAT 1500
QY 1501 GAAGATTTTATTCGCAATTCGATCATATTTTCTCCCGGAACACCTTGTGGAATGGT 1560
DB 1501 GAAGATTTTATTCGCAATTCGATCATATTTTCTCCCGGAACACCTTGTGGAATGGT 1560
QY 1561 TTAACAGCAGACCGCGCGCAGAGATGGATTTACTCCCGGCACACCTCTCGCTGTAGGG 1620

Db 1561 TTAACAGCACAAAGCGCGCAGAGATGGATTACTCCCGGCGACACCTCTCGCTGTAGGG 1620
Qy 1621 TTAATGTATGCTACAGCTGGTGGCATCGGTACGGTTGGCGTAGAAGGTGGAGCGCTGAAC 1680
Db 1621 TTAATGTATGCTACAGCTGGTGGCATCGGTACGGTTGGCGTAGAAGGTGGAGCGCTGAAC 1680
Qy 1681 AATCTCGGCTATGTTTTCGGCACTTCTTCATGCACATGCGCATCTACCACTTCTCCCTCG 1740
Db 1681 AATCTCGGCTATGTTTTCGGCACTTCTTCATGCACATGCGCATCTACCACTTCTCCCTCG 1740
Qy 1741 TTTGTACCGGGTGTCTGGGGCGGTATTACAGTGCAGATGGTTCCAGGTCTGTGGTTAGTT 1800
Db 1741 TTTGTACCGGGTGTCTGGGGCGGTATTACAGTGCAGATGGTTCCAGGTCTGTGGTTAGTT 1800
Qy 1801 GAAGCGGGCAAGTGTCTCGGGGAGCAGCTATTGACCACTACTTGAATTCATCCCGGCT 1860
Db 1801 GAAGCGGGCAAGTGTCTCGGGGAGCAGCTATTGACCACTACTTGAATTCATCCCGGCT 1860
Qy 1861 GTTGAAGAGCTCGGAAATGCGCAACATCAGATGCTGTGCGCCCTCGCGAAAGGGCTA 1920
Db 1861 GTTGAAGAGCTCGGAAATGCGCAACATCAGATGCTGTGCGCCCTCGCGAAAGGGCTA 1920
Qy 1921 GATCGAATCCTCGAAAAACGGCGCAACCATCAGATGCTGTGCGCCCTCGCGAAAGGGCTA 1980
Db 1921 GATCGAATCCTCGAAAAACGGCGCAACCATCAGATGCTGTGCGCCCTCGCGAAAGGGCTA 1980
Qy 1981 CACGTGGTCCGGAATTTCTCGGAAATCGCGCCCTCGCGAGATCCTCATGCCAGAGCG 2040
Db 1981 CACGTGGTCCGGAATTTCTCGGAAATCGCGCCCTCGCGAGATCCTCATGCCAGAGCG 2040
Qy 2041 GTAATTTGTGGCTGGGTATGAGGCGAGATCTGGATAATTTACTCGCCCTTGATATCGCT 2100
Db 2041 GTAATTTGTGGCTGGGTATGAGGCGAGATCTGGATAATTTACTCGCCCTTGATATCGCT 2100
Qy 2101 GGATTATGCGGAATTTGTTATGCTGTGCGCCAAATTTCTGACGCTCAAAACAGCGAGGA 2160
Db 2101 GGATTATGCGGAATTTGTTATGCTGTGCGCCAAATTTCTGACGCTCAAAACAGCGAGGA 2160
Qy 2161 GTAGTGAGTAAATATCGTTATTAGCGCGGTGGCGGAGCATCCACTGGTACGACAA 2220
Db 2161 GTAGTGAGTAAATATCGTTATTAGCGCGGTGGCGGAGCATCCACTGGTACGACAA 2220
Qy 2221 ATTCTGGCAGATACCTGGGTATTCCGGTCAATTACCGCAATGCTGCGAAGCGTTTAA 2280
Db 2221 ATTCTGGCAGATACCTGGGTATTCCGGTCAATTACCGCAATGCTGCGAAGCGTTTAA 2280
Qy 2281 TTAGGCTCGGCCATTTCTGGTGTCTGCGCGGAAATATTGACACTTCTGTTGGCGAAGCG 2340
Db 2281 TTAGGCTCGGCCATTTCTGGTGTCTGCGCGGAAATATTGACACTTCTGTTGGCGAAGCG 2340
Qy 2341 ATGCAACAAATCACCCATGTGGATAAATATTATTCCGCAAGACGCTATCAATCTCTT 2400
Db 2341 ATGCAACAAATCACCCATGTGGATAAATATTATTCCGCAAGACGCTATCAATCTCTT 2400
Qy 2401 CATCATGTCGATATGAGCGCTATAGCAGTTGCGAGCATACGCAAAATTTACTCGAGAC 2460
Db 2401 CATCATGTCGATATGAGCGCTATAGCAGTTGCGAGCATACGCAAAATTTACTCGAGAC 2460
Qy 2461 TAAATTAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATACGACTCACTCCGGTA 2520
Db 2461 TAAATTAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATACGACTCACTCCGGTA 2520
Qy 2521 ATATCCCGAGTGCATTTCTATCGACCTTAAACAGAGTTTATATGTCGCAAAATAATA 2580
Db 2521 ATATCCCGAGTGCATTTCTATCGACCTTAAACAGAGTTTATATGTCGCAAAATAATA 2580
Qy 2581 AACAGTGGTGGTTGGCACTGCACTGTGATATGGGATATATCGGCATTCGCGATTTTA 2640
Db 2581 AACAGTGGTGGTTGGCACTGCACTGTGATATGGGATATATCGGCATTCGCGATTTTA 2640
Qy 2641 TGACTGGTGATGGTTTCGAACTCGCATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700
Db 2641 TGACTGGTGATGGTTTCGAACTCGCATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700

Qy 2701 CACCTGCGGAAGCCTCTTTTGGCTTTACGCTCTACGCGCTGGCGCTGCCCTTTTCGCGCT 2760
Db 2701 CACCTGCGGAAGCCTCTTTTGGCTTTACGCTCTACGCGCTGGCGCTGCCCTTTTCGCGCT 2760
Qy 2761 GGGTTTCTGGGTAGTAGGGAAATCATCAGCCGCAAAAACCATCTGATGTTGTTTTG 2820
Db 2761 GGGTTTCTGGGTAGTAGGGAAATCATCAGCCGCAAAAACCATCTGATGTTGTTTTG 2820
Qy 2821 TCCTATGGTGGCTATTCCATGTTCTGTTCTGGTCTTTGGATTAGGACAGGCAAACTATG 2880
Db 2821 TCCTATGGTGGCTATTCCATGTTCTGTTCTGTTCTTTGGATTAGGACAGGCAAACTATG 2880
Qy 2881 GGTAAATCTGCTGTTTACGGATTCGGTTTACGGTATCGGCTATTTCTCTACTCAT 2940
Db 2881 GGTAAATCTGCTGTTTACGGATTCGGTTTACGGTATCGGCTATTTCTCTACTCAT 2940
Qy 2941 TTATCGTCTGTTTATTTATTAACAGTGCAGTGCAGAAAATCCAGTTCTGCTCGGTGGT 3000
Db 2941 TTATCGTCTGTTTATTTATTAACAGTGCAGTGCAGAAAATCCAGTTCTGCTCGGTGGT 3000
Qy 3001 ACTGGCGGTATATTCACTAGGATCGGTGTTGCTGGCAGTTATATTTCCAGTTCACGA 3060
Db 3001 ACTGGCGGTATATTCACTAGGATCGGTGTTGCTGGCAGTTATATTTCCAGTTCACGA 3060
Qy 3061 TACCGATTATGGGTGAAATGGGAACCTTATGGCTGGCAGTGGCGTTCTGCTTTGCTGGCG 3120
Db 3061 TACCGATTATGGGTGAAATGGGAACCTTATGGCTGGCAGTGGCGTTCTGCTTTGCTGGCG 3120
Qy 3121 GTGTCAATGCCATGATCTCTTCCGTCATGTTAAACGCCCTGGACATATGCAATAATTA 3180
Db 3121 GTGTCAATGCCATGATCTCTTCCGTCATGTTAAACGCCCTGGACATATGCAATAATTA 3180
Qy 3181 CTCGCCGTGAGAAATTTGAGAAATTAAGTCGGGCACTAACTTTACTTTATACCAACGCA 3240
Db 3181 CTCGCCGTGAGAAATTTGAGAAATTAAGTCGGGCACTAACTTTACTTTATACCAACGCA 3240
Qy 3241 ATATTTTCTCTCCAGTATGTGCGCATATATAAATACCTTATCGTTATTCGGTTTTCGCG 3300
Db 3241 ATATTTTCTCTCCAGTATGTGCGCATATATAAATACCTTATCGTTATTCGGTTTTCGCG 3300
Qy 3301 TCATTAATGCCAATGATGTTTGGATGAACCTGGGATTCACCACTCTGAATGCTGCGAG 3360
Db 3301 TCATTAATGCCAATGATGTTTGGATGAACCTGGGATTCACCACTCTGAATGCTGCGAG 3360
Qy 3361 TCTGGCGGCAATTTTCTTTACCACTATTTCTCTAATATTTTGGGGATTCGTCAG 3420
Db 3361 TCTGGCGGCAATTTTCTTTACCACTATTTCTCTAATATTTTGGGGATTCGTCAG 3420
Qy 3421 AAAAAATGGCTGGATGCGTGTATTTCGCTGGTGTGTTGCTCGGGATGGCAGATCA 3480
Db 3421 AAAAAATGGCTGGATGCGTGTATTTCGCTGGTGTGTTGCTCGGGATGGCAGATCA 3480
Qy 3481 GTTTAGCTGTTTACTACATGCCCAATACTTTGGTGCATTTGCTGCGGATGATTC 3540
Db 3481 GTTTAGCTGTTTACTACATGCCCAATACTTTGGTGCATTTGCTGCGGATGATTC 3540
Qy 3541 CGCGGATTCCTGCGCACTTTTGTGCTGCAATTTGTTGCGGATGGCGCTGCTTCCCGG 3600
Db 3541 CGCGGATTCCTGCGCACTTTTGTGCTGCAATTTGTTGCGGATGGCGCTGCTTCCCGG 3600
Qy 3601 CACTGGAACCAAAACAGGTGCTGCAATCTCGGTTCACCACTCTCTCGCGGTATGT 3660
Db 3601 CACTGGAACCAAAACAGGTGCTGCAATCTCGGTTCACCACTCTCTCGCGGTATGT 3660
Qy 3661 CTAACTTCCTGGCTCGGCAATTCGGTGTGTTTACCGTGGTTAGCACTATCGGTG 3720
Db 3661 CTAACTTCCTGGCTCGGCAATTCGGTGTGTTTACCGTGGTTAGCACTATCGGTG 3720
Qy 3721 TGGTCAATTCCTATACAGCATGCTATCTATTTGCGCTTTGCTCTTGGCAATTCATTCGCG 3780
Db 3721 TGGTCAATTCCTATACAGCATGCTATCTATTTGCGCTTTGCTCTTGGCAATTCATTCGCG 3780

The invention relates human kinases (PKIN) and their corresponding nucleic acid sequences. PKIN and its DNA are useful for diagnosing, treating and preventing cancer, an immune system disorder (e.g., acquired immune deficiency syndrome (AIDS), Addison's disease, allergy asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders

Db 508 TTTTTTGGATAGCAGTTTACCCCATTTACCACTAACAGGAGGGATTCCTCCATCGAAA 567
Qy 1179 TATCATTTGTGGATGATACCGCGCCACCGAACAACAGCAGACCAATCAATGCCACTCA 1238
Db 568 CGTCATCATGTGGTGGACCATCGACAGTCACTAGTCAAGTTACAGAGATCAATGACACCA 627
Qy 1239 CCATCCGGTGTGAACACTACGTGGGTGTAATTTTCGGCTGAAATGGAACACCGCAAAAT 1298
Db 628 GCACAGTGTCTCCAGTACGTGGGGGGGTGATGCTGTGGAATGCGAGGCCCGCAAACT 687
Qy 1299 TCTCTGGCTGAAGAAATATGCCAGAGAT---CTACCAACGTGCCGGACAAATTTTTCGA 1355
Db 688 TCTGTGGCTGAAGAGAACTTGAAGAGATTTGCTGGGATGAAGCGGGACATTTCTTTTGA 747
Qy 1356 TCTGCGCGATTTCTGACCTGGCGGCTACCGGTGATTTAGCGCTTCAGTATGCACATGT 1415
Db 748 TCTCCACAGCTTCTATCGTGAAGCAACAGGTGTACAGCAGCGTCTCTGCTGCCCT 807
Qy 1416 TACCTGTAATGGACGTGGCTGGCACATGAAATCGCTGGATCCAGATTTATTTCCGCAC 1475
Db 808 GGTGTGAAGTGGACATATTACGCA---GAGAAAGCTGGGACGACAGTTCCTGGAATA 864
Qy 1476 CATTTGGCTTGCAGAGATTAGCGGATGAAGATTTTATTCGCATTTGTCATCATATTTTTC 1535
Db 865 GATTTGGTTTGAAGACTTTTGTTCAGATAATTACAGCAAAATAGGAAACCAAGTGTCTACC 924
Qy 1536 TCCCGGAACACTTCTGGAATGPTTAACAGCAACGCCGCGCAGAGATGGGATTACT 1595
Db 925 TCTTGGAGCTTCTCTTGGAAATGGCTCACACGAGGACAGACAGTGTGGCTTCT 984
Qy 1596 CCCCCGACACTGTGCTGTAGGGTTAATTGATCTCACGCTGTGGCATCGGTACGGT 1655
Db 985 CCTGGGATTGCGGTGCGAGCTTCACTATTGATGCCCATGACGAGGACTAGGAGTGTAT 1044
Qy 1656 TGGCTAGAAGTGGAGCGC-----TGAAACAATCT 1685
Db 1045 TGGGCGAGATGTGAGAGGGCACCGGCTCATCTGTGAGGGGACGCCAAGTGCACGCT 1104
Qy 1686 CCGTATGTTTTGCGCACTTCTTATCACCATGCATCTACCACTTCTCCCTCGTTTGT 1745
Db 1105 GGTCTCATCTGTGAAGCTTCTTGTGCATGGGGATCAGCAAGACCCGATTTTGT 1164
Qy 1746 ACCGGGTGTCTGGGGCGGTATTAACAGTGCAGTGTCCAGTCTGTGTAGTTGAAGG 1805
Db 1165 ACCAGCGCTGGGGCGCTTATTTCTCAGCCATGCTACCTGGGTCTGCTGAATGAAGG 1224
Qy 1806 CGGGCAAGTGTCTGGGAGCAGCTATTGACCAGCTACTTGAATTCATCCCGCTGTTGA 1865
Db 1225 TGGTCAGAGCGTTACTGAAATTTGATAGACCACATGGTACAAAGGCCATGCTGCTTTTCC 1284
Qy 1866 AGAAGCTCGGAAATGGCACAACGTGTGAATCAGCCCTCCCGCTGCTGCTGCTGATCG 1925
Db 1285 AGAAGTACAAGTAAGGCCACAGCCAGATGCCAGAGTATATATGAT---ATTGAACAG 1341
Qy 1926 AATCCTCGAAACCGGCGCAACCATCAGATGCTGCTGCCCTGGCGAAAGGCTACAGCT 1985
Db 1342 TCACCTGGATCTGATTAAGAAGGCTCAGCCTGTGGGTTTCCCTACTGTTGATTTACATGT 1401
Qy 1986 GGTGCGGGAATTTCTCGGAATTCGGCCCGCCCTTCCAGATCCTCATGCCAGAGCGGTAT 2045
Db 1402 TTGGCGGATTTCCCATGCAACCGGCTCTCCCTTAGCAGATCTGACACTAAAGGGCATGT 1461
Qy 2046 TTGTGCGCTGGGTATGGAGCGAGATCTGGATAATTTACTCGCCTTGTATATCGCTGGATT 2105
Db 1462 CACCGGATTAAGTGTCTCAGGACCTTGATGATCTTGCCATTTCTACCTGGCCACAGT 1521
Qy 2106 ATCGGGAATGGTATGCTGCGCAAAATTCGACGCTCAACAGCGCGAGGAGTGTAGT 2165
Db 1522 TCAAGCCATTTGCTTTGGGAGCTCGCTTCTATTATAGAAGCCATGGAGGACGAGGCACTC 1581
Qy 2166 GAGTAAAAATTCGTTTATAGCGGGGTCGCGGGGAGCATCCAGTGGTACGCAAAATCT 2225

Db 1582 AATCAGTACTCTTTTCCCTATGTGGAGGCTCAGCAAGAAATCCCTTTTGTGCAAAATCA 1641
Qy 2226 GGCAGATACCTCGCGTATTTCCGGTCAATACCAACCAATGCTCGCAACCGGTTTATTAGG 2285
Db 1642 TCGGACATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1701
Qy 2286 CTCGGCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2345
Db 1702 TCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1761
Qy 2346 ACAATTCACCCATGCGATAAATATATATCCCAAGAACGCTATCAATCTCTTCATCA 2405
Db 1762 AAAAATGAGCAAGTTGGAAAGTTGTGTTCCCGAGACTACAGATAAAAAATACTATGA 1821
Qy 2406 TCGTGCATATGAGGCTATAAGCAGTTG 2433
Db 1822 TAAGAATACCAAGTATTCCTGAAGCTG 1849
RESULT 5
AAS03016
ID AAS03016 standard; cDNA; 1991 BP.
XX AAS03016;
AC AAS03016;
XX XX
DT 29-AUG-2001 (first entry)
XX Human diagnostic and therapeutic (dithp) cDNA sequence #5.
XX Human diagnostic and therapeutic molecule; dithp; gene therapy;
KW thalassemia; cardiovascular disorder; cell proliferative disorder;
KW cancer; neurodegenerative disorder; autoimmune disorder; enzyme;
KW infectious disorder; inflammatory disorder; developmental disorder;
KW Incyte ID number 05377811dec; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO200121836-A2.
XX 29-MAR-2001.
XX 19-SEP-2000; 2000WO-US25643.
XX 23-SEP-1999; 99US-0155760.
PR 24-SEP-1999; 99US-0155939.
PR 24-SEP-1999; 99US-0156294.
PR 28-SEP-1999; 99US-0156565.
PR 28-SEP-1999; 99US-0156624.
PR 28-SEP-1999; 99US-0156625.
PR 24-NOV-1999; 99US-0167410.
PR 24-NOV-1999; 99US-0167453.
PR 24-NOV-1999; 99US-0167517.
PR 24-NOV-1999; 99US-0167520.
PR 24-NOV-1999; 99US-0167542.
PR 29-NOV-1999; 99US-0167943.
PR 29-NOV-1999; 99US-0167945.
PR 30-NOV-1999; 99US-0168197.
PR 30-NOV-1999; 99US-0168265.
PR 30-NOV-1999; 99US-0168429.
PR 30-NOV-1999; 99US-0168432.
PR 01-DEC-1999; 99US-0168468.
PR 01-DEC-1999; 99US-0168599.
XX (INCY-) INCYTE GENOMICS INC.
XX Hodgson DM, Lincoln SE, Russo PD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HU, Rosen BH, Shah P, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR;
PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK;
PI Amshey S, Fong WT;
XX WPI; 2001-281607/29.
XX

PT Novel diagnostic and therapeutic polynucleotides, used in disease
PT diagnosis and for gene therapy of conditions such as cancer and
PT thalassemia

PS Claim 1; Page 253-254; 299pp; English.

XX The present sequence for human diagnostic and therapeutic (dithp) cDNA
CC sequence #5 is 1 of 71 (AA503012-AA503082) novel sequences described
CC in the invention. The present sequence (Incye ID No: 05377811dec)
CC encodes an enzyme molecule. The dithp polynucleotides may be used
CC to diagnose a condition disease or disorder associated with human
CC molecules. They can be used to identify the presence of similar
CC nucleic acids. Dithp polynucleotides may be used to generate hybridisation
CC probes for use in chromosomal mapping. Polypeptides (dithp) encoded by
CC dithp are used to screen for molecules which bind to them and modulate
CC their activity. Dithp polynucleotides can be used for gene therapy of
CC disorders such as severe combined immunodeficiency syndrome (SCID),
CC cystic fibrosis, thalassemia, haemophilia resulting from Factor VIII
CC or IX deficiencies, cardiovascular disorders, e.g familial
CC hypercholesterolaemia (FH), cell proliferative disorders e.g. cancers,
CC neurodegenerative disorders, autoimmune/inflammatory disorders,
CC infectious disorders and developmental disorders. The antibodies can be
CC used to analyse protein expression levels.

XX Sequence 1991 BP; 468 A; 467 C; 534 G; 516 T; 6 other;

XX Query Match 7.6%; Score 300.8; DB 22; Length 1991;
XX Best Local Similarity 52.0%; Pred. No. 1.3e-84;
XX Matches 826; Conservative 0; Mismatches 717; Indels 45; Gaps 5;

QY 879 TATTGTGTGTGATGGGATCAGGAGTGTGCGCGCGGGATTTTGTATCTCAACGGATC 938
DB 227 TGTGGGTGTGGAGTTTGGAAACAGGCACTGTCGTCAGCTGTGGTGACACAGAGTGGGT 286
QY 939 TCTGCTATCCCATGCCACAGAAAAATCAGACTACGCGCGGCGAGCGGAGCGCGTGA 998
DB 287 CCTGTTGGCTTTTGGACAGCAGCAATTAAGAAATGGGAGCCCCAGTTCAACCCACCATGA 346
QY 999 ACAGTCAGCCAGGAGATCTGCGAGCGGTCTGTTTCATGTATTCGTAATTCGCGTCACTCT 1058
DB 347 GCGATCCTCCGAGGACATCTGGGCTGCGTGTGTGTGTCACAAAGAAAGT-----TGT 400
QY 1059 GCGAGAGCTTTGTGCAACAAGTGTGCGCAGGATCGGTTTGTATGCCACCTGTCTCTGTT 1118
DB 401 ACAAGGGATTTGATTTAAACCAAAATTCGAGGACTTGGGTTTGTATGCCACGTCTCTCTGT 460
QY 1119 GGTACTGGATAAAACGGTGTCCATTGCTGTCAGCGCGGAGGAGATGCAAGCAAAA 1178
DB 461 TGTTTTGGATAGCAGTTTCAACCATACAGTCAACCAAGTAAAGGGATTCGCATCGAAA 520
QY 1179 TATCATTTGTGTGGATGATCACCGCGCCACCGAACAGCAGACGCAATCAATGCCACTCA 1238
DB 521 CGTCATCATGTGCTGGACCATCGAGCATCAGTCAAGTTAAACAGGATCAATGAGACCAA 580
QY 1239 CCATCCGGTGTGAACTAGCTGGTGGTAAATTTCCGCTGAAATGGGAACACCGGAAAT 1298
DB 581 GCACAGTGTCTCCAGTACGTGCGGGGGGTGATGTGTGGAATGACGGCCCCGAAACT 640
QY 1299 TCTCTGGCTGAAGAAATATGCCAGAGAT---CTACGAACGTGCGCGACAAATTTTTCGA 1355
DB 641 TCTGTGGCTGAAGAGAACTTGAGAGAGATTTGTCTGGGATAGCGGGAGACATTTCTTTGA 700
QY 1356 TCTGGCCGATTTTCTGACCTGGCGGCTACCGGTGATTTAGCGGTTTCAGTATGCACTGT 1415
DB 701 TCTCCGGGACTTCTTATCTGTGGAAGCAACAGGTGTCAGCAGCGGTCTCTGCTCCCT 760
QY 1416 TACCTGTAATGGACGTGCTGGCAGATGAAATCCCTGGGATCCAGATTTATTCGGCAC 1475
DB 761 GGTGTGTAAGTGSACATATTTCAGCA---GAGAAAGCTGGGACGACAGTTTCTGGAAT 817
QY 1476 CATTGGCTTGGACAGTATAGCGGATGAAGATTTTATTCGATTTGGTCATCATATTCTTTC 1535
DB 818 GATTGGTTTGGAGACTTTGTTGCAGATAATTACACCAAAATAGGAAACCAAGTGTCTACC 877

QY 1536 TCCCGGAACACCTTGTGAAATGTTTAAACAGCACAAAGCCGCGGACAGATGGGATTACT 1595
DB 878 TCCTGGAGCTTCTCTTGGAAATGGGCTCACACAGAGGAGCAAGAGACTTGGCCTCT 937
QY 1596 CCGCGCACACCTGTGCTGTAGGGTTAAATTGATCTCAGCTGGTGGATCCGTACGGT 1655
DB 938 CCTTGGATTTGGCGTGCAGCTTCACTCATTTGATGCCCATGCGAGGAGACTTGGAGTAT 997
QY 1656 TGGCGTAGAAGTGGAGCGC-----TGAAACAATCT 1685
DB 998 TGGGCGAGATGTGAGAGGCGACGCGCTCATCTGTGAGGGCAGCCAGTACGTCACGGCT 1057
QY 1686 CCGCTATGTTTTCGGCACTTCTTTCATGCACCATGGCATCTACCACTTCTCCCTCGTTGT 1745
DB 1058 GCGTGTCTATCTGTGGAACCTCTTCTTGTACATGGGATCAGCAAAAGCCGATTTTGT 1117
QY 1746 ACCGGTGTCTGGGGCGCTATTTACAGTCCGATGGTTCAGGCTCTGTGTAGTCTCAAGG 1805
DB 1118 ACCAGCGCTCTGGGGCGCTTATTTCTACGCCATGGTACCTGGGTCTGCTGTAATGAAG 1177
QY 1806 CCGGCAAAAGTGTGCGGGAGCAGCTATTGACCAAGCTACTTGTATTCATCCGCGCTTGA 1865
DB 1178 TGGTCAGACGCTTACTTGGAAATTTGATAGCCACATGTTACAGGCCATGCTCTTTC 1237
QY 1866 AGAAGCTCGGAAATGGCAACAGTGTGAATCAGCCCCCTCCCGCTCTGGCTTGTCTGATCG 1925
DB 1238 AGAATACAAAGTAAAGGCGCACAGCCAGATGCCAGATATATATGCAAT---ATTGAACAG 1294
QY 1926 AATCTCTGAAAAAAGCGCGCAACCATCAGATGCTGCGCCCTGGCGAAGGGCTACAGT 1985
DB 1295 TCACCTGGATCTGATTAAGAAGCTCAGCGCTTGGGTTTCTTACTGTTGATTTACATGT 1354
QY 1986 GGTGCGGAAATTTCTCGGAAATCGCGCCCTTCGCAGATCTCCATGTCACAGAGCGGTAA 2045
DB 1355 TTGCGCAGATTTCCATGGCAACCGGTCTCCCTTAGCAGATCTGCACACTAAAGGGCATGG 1414
QY 2046 TTGTGCGCTGGGTATGGAGCGAGATCTGATAATTTACTGCGCTTGTATATCTGCTGAT 2105
DB 1415 CACCGGATTTGAAACTGTCTCAGGACCTTGATGATCTTGCCATTTCTACCTGGCCACAGT 1474
QY 2106 ATGCGGAATTTGTTGTTCTGGCCAAATTTCTGAGCTCAAAACAGCGCAGGAGTAGT 2165
DB 1475 TCAAGCCATTGCTTTGGGACTCGCTTCATATAGAGGCCATGGAGGACAGCGGCACCT 1534
QY 2166 GAGTAAATAATCTGTTATTAGCGCGGTGCGGGCAGCATCCACTGTACGACAAAATCT 2225
DB 1535 AATCAGTACTCTTTTCTATGTGGAGCCCTCAGCAGAAATCCCTTTTGTGCAATGCA 1594
QY 2226 GGCAGATACCTCGGTTATTTCCGGTCAATTACCAGCAATGCTGCGAAACCGGTTTATTAGG 2285
DB 1595 TGGGACATTTACTGGCATGCTGTGGTCTGTGCGAAGAGTGGAGTCCGTTCTTGTGGG 1654
QY 2286 CTCGGCCATTTGTTGCTGTCGCCGGAATATTGACACCTTCTGTTGGGCAAGCGATGCA 2345
DB 1655 TGTCTGTTCTGGGTGCGCTGTCCTCAGGGGATTTGCTTCTGTACAGGAACAATGGC 1714
QY 2346 ACAATTCACCCATGTGGATAAATAATTATTATCGCAAGAACGCTATCAATCTCTTCATCA 2405
DB 1715 AAAAAAGCAAAAGTTGGGAAAGTGTGTTCCGAGACTACAGGATAAAAAATCTATGA 1774
QY 2406 TCGTCGATGAGGCGCTATAAGCAGTTG 2433
DB 1775 TAAGAAATACCAAGTATTCCTGAAGCTG 1802

RESULT 6
ABL26431
ID ABL26431 standard; DNA; 2602 BP.
XX
AC ABL26431;
XX
DT 26-MAR-2002 (first entry)

Db 249 CCTGTGGCTTTTGAGACAGCAATTAAGATTGGAGCCCGCAGTTCAACCAACCATGA 308
 QY 999 ACAGTCCAGCAGAGATCTGCGACGGCGTGTCTGATGTAATGCGCTCACTCT 1058
 Db 309 GCAATCCCTCCAGGACATCTGGCGTGGCTGTGTCTTCACAAAGAAAGT-----TGT 362
 QY 1059 GGCAGACGTTTGTGCACAAAGTGTGGCAGCATCGTTTGTGATGCGACCTGTTCTGTGT 1118
 Db 363 ACAAGGATGATTTAAACCAAAATTCAGAGACTTGGGTTTGTATGCCAGGTGTTCTGTGT 422
 QY 1119 GGTACTGATTAACAGGATGATTCCTCTGTACGCCGAGAGAGATGCAAGCAAAA 1178
 Db 423 TGTTTTGGATAGCAGTTTCAACCATTCACCAACGAGAGGGGATTCCTCATCGAAA 482
 QY 1179 TATCATTTGTGATGATGATCAACCGCGCACCAAGCAAGCAGAGCATCAATGCGCATCA 1238
 Db 483 CGTCACTCTGTGTGCTGCGACCATTCAGACAGTCAATCAATTAACAGATCAATGAGACCA 542
 QY 1239 CCATCCGCTGCTGAACACTACGTCGGTGG 1265
 Db 543 GCACAGTGTCTCCAGTACGTCGGGGG 569
 RESULT 11
 AAH04701
 ID AAH04701 standard; cDNA; 795 BP.
 AC AAH04701;
 DT 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:1536.
 DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 PN EPI074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 XX Oda T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 1; SEQ ID 1536; 2537pp + CD ROW; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 SQ Sequence 795 BP; 174 A; 206 C; 224 G; 188 T; 3 other;
 Query Match 2.6%; Score 104.2; DB 22; Length 795;
 Best Local Similarity 57.7%; Pred. No. 5.6e-22;
 Matches 241; Conservative 0; Mismatches 171; Indels 6; Gaps 3;
 QY 1160 AAGGATGCAAGCAAAATATCATTTGTGTGATGATCACCGCCACCGCAAGCAG 1219
 Db 380 AAGGGATTTCCATCGAAGACGATCATCATGTGCTGGACCATGACAGTCAAGTTA 439
 QY 1220 AGCGAATCAATGCCACATCCGATCCGCTGTAAGTACGTCGGTGAATTTGCGCTG 1279
 Db 440 ACAGGATCAATGAGACCAAGCAGATGCTCCAGTACGTCGGGGGATGATGCTGTGG 499
 QY 1280 AATGGAACACCGAAATTTCTGTGCTGGAAGAAATATGCAAGAT---CTACGAAAC 1336
 Db 500 AATGCAAGGCCCGCAAACTTGTGTGCTGGAAGAAACATTTGCTGGGATA 559
 QY 1337 GTGCCGCAATTTTGCATGCGCCGATTTTCTGACCTGGCGGCGTACCGGTGATTAG 1396
 Db 560 AGCGGCGACATTTCTTTATATCTCCCGACTTCTTATCGTGAAGCAACAGGTGCAACAG 619
 QY 1397 CGCGTTCAATGATGACATGTTTACCTGTAATGAGACGTGCGCATGAAATTCGCTGG 1456
 Db 620 CACGGGCTCTGCTGCTCCCTGGGTGTAA--GTGACATATTCACACAGAAAGCTGGG 677
 QY 1457 ATCCAGATTATTTCCGACACATTTGGCTTGCAGAGTATGCGGATTAAGATTTTATTCGA 1516
 Db 678 ACGACAGTTTCTGGAATAATGATGATGGAAGACTTTGTGTGCAG-ATAATTTCACAAAA 736
 QY 1517 TTGTCATCATATTTGTTCTCCCGAACAACCTTGTGAAGTGAAGTGAACAGCAGAC 1574
 Db 737 TAGGAACCAAGTCTACTTCTGAGACTTCTTGTGAAGTGGCTNACACCAAGANCC 794
 RESULT 12
 ID ABBN60111 standard; cDNA; 677 BP.
 AC ABBN60111;
 DT 28-JUN-2002 (first entry)
 XX Human cancer related polynucleotide SEQ ID NO 78.
 DE Human; cytostatic; gene expression; gene mapping; tissue profiling;
 KW gene therapy; cancer; tumour; gene; ss.
 OS Homo sapiens.
 XX WO200214500-A2;
 PM 21-FEB-2002.
 PD 16-AUG-2001; 2001WO-US25840.
 PR 16-AUG-2000; 2000US-226326P.
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

XX Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
 PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
 XX WPI: 2002-241905/29.
 XX
 PT New nucleic acid for producing a polypeptide, detecting differentially
 PT expressed genes correlated with a cancerous state of a mammalian cell,
 PT and inhibiting tumor growth -
 XX
 PS Claim 1: SEQ ID NO 78; 883bp + Sequence Listing: English.
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
 CC with cytosolic activity. The polynucleotide is used to produce a
 CC polypeptide, to detect differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell and to inhibit tumor growth. The
 CC polynucleotide is used as a probe in mapping and tissue profiling. The
 CC encoded polypeptide and antibodies to the polypeptide can also be used
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
 CC gene therapy.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 677 BP; 157 A; 164 C; 184 G; 172 T; 0 other;

Query Match 2.3%; Score 90.4; DB 24; Length 677;
 Best Local Similarity 49.4%; Pred. No. 1.3e-17;

Matches 321; Conservative 0; Mismatches 296; Indels 33; Gaps 2;

OY 1525 CATATTGTTTCCCGGACACCTTGGAATGTTTAAACACAAACCGCGGACAG 1584
 DB 1 CAAGTCTACCTCTGAGAGCTTCTTGGAAATGGCTCACACAGAGCAGCAAGAGAC 60
 OY 1585 ATGGAGTATCCCGGACACCTCTGCTGTAGGGTTATGATGCTCAGCGTGGC 1644
 DB 61 CTGGGCTTCTCCCTGGGATGGCGTGGCAGCTTCACTATGATCCCATGACGAGGA 120
 OY 1645 ATCGGTACGCTGGCTAGAGGTGAGCGCTAACAACTCGCGTATG----- 1693
 DB 121 CTAGAGTATGGGGGAGATGAGAGGGGACAGGCTCTGTAAGGGGACAGCAGT 180
 OY 1694 -----TTTCCGACATCTTCATGACCATGAGCATCTACCACTTC 1734
 DB 181 ACGTACGCGCTGCTCATCTGTGAGACGCTCTTCTTGACATGGGGACACAAAGAC 240
 OY 1735 CCCTGCTTTTACCGGGGTCTGGGGGCGGTATTAACAGTGCAGTGGTCCAGGTCTGTG 1794
 DB 241 CCGATTTTGTACAGGCGCTGGGGGCTTATTTCTCAGCCATGGTACTGGGTTCTG 300
 OY 1795 TTAGTTGAAGGCGGCAAGTCTGGGAGACAGCTATTGACCACTTGTGATTTCCAT 1854
 DB 301 CTGAATGAAGGTGTCAGAGCGTTACTGGAATAATGATAGCACATGTGACAAAGCCAT 360
 OY 1855 CCGGCTGTGAAGACGTGCGGAATGCGCAACAGTGTGATGATGAGCCCTCCCGCTG 1914
 DB 361 GCTGCTTTTCCAGAACTTACAGTAAGGCGACAGCCAGATGCCAGATATATATGAT-- 418
 OY 1915 CTTCGTGATGATCTTCGAAAAAAGCGGCAACATCATGATCTCTCCCGCGGAGAA 1974
 DB 419 -ATTGAACAGTACCTGTGATGATTAAGAAAGCTCAGCCTGTGGGTTTCTTACTGTT 477
 OY 1975 GGGCTACACGTGGTCCGGAATTTCTCGGAATGCGGCCCTTCGCAATCTCATGCG 2034
 DB 478 GATTTACATGTTTGGCAGATTTCCATGGCAACCGGTCTCCCTTACAGATCTGACACTA 537
 OY 2035 AAGAGCGTATTTTGGCTGCGGTATGAGAGCAGATCTGATATTTACTCGCTTGTAT 2094
 DB 538 AAGGCGATGATCAGCGATTTGAACGTCTCAGGACCTGATCATCTTGCATTTCTCTAC 597
 OY 2095 ATCGGTGATTAATGCGGAATTTGTTATGTCGCGCAATTTCTGACGG 2144
 DB 598 CTGGCAGACGTTCAAGCCATTGCTTTGGGAGCTCGCTTCATATATAGAGC 647

RESULT 13
 AAC02752
 ID AAC02752 standard; cDNA; 409 BP.

XX AAC02752;
 AC
 XX 06-OCT-2000 (first entry)
 DT
 XX

DE Human secreted protein 5' EST, SEQ ID NO: 2750.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.
 XX P-PSDB: AAC02746.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 2750; 721bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 CC

SO Sequence 409 BP; 103 A; 89 C; 115 G; 102 T; 0 other;

Query Match 2.2%; Score 86.2; DB 21; Length 409;
 Best Local Similarity 56.8%; Pred. No. 2e-16;

Matches 179; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

OY 1329 CTACGACGTCGCGCAATTTTTCATCTGCGGCAATTTCTGACCTGGCGGCTACCGG 1388
 DB 93 CTGGGATGAAGCGCGGACATTTCTTGTATCTCCCGGACTCTTATCTGGAAGCAACAG 152
 OY 1389 TGATTTAGCGCGTTGAGTACGTACTGTAAATGAGCGTGGCTGCAATGAAAA 1448
 DB 153 TGTCAACAGCAGGCTCTCTGCTCCCTGGTGTAAAGTGACATATTCACGA---GAGA 209
 OY 1449 TCGCTGGATCCAGATTATTTCCGACCAATTGGCTTGACAGTTAGCGGATGAAGATT 1508
 DB 210 AGGCTGGAGACAGATTTCTGGAATAATGATTGGTTGGAAGACTTGTGACAGATAATTA 269
 OY 1509 TATTCGATTTGATCATATTTCTTCTCCGGAACACTTGTGGAATAATGTTAACACG 1568

Db 270 CAGCAAAATAGCAACCAAGTGTACTCTCTGGAGCTTCTTGAAATGGGCTCAGACC 329
Oy 1569 ACAACCGCGGAGAGATGATATCTCCCGGACACCTGTCGTTAGGTTAATGA 1628
Db 330 AGAGGACAGCAAGACCTTGCTCTCCCTGGATTTGGCGATGCAAGCTTCACATATGA 389
Oy 1629 TGCTCAGCGCTGTGG 1643
Db 390 TGCCCATGCGAGGAGG 404

RESULT 14
AAH07400
ID AAH07400 standard; cDNA: 613 BP.
XX
AC AAH07400;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4235.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Mshikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 4235; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 3'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 613 BP; 146 A; 140 C; 164 G; 158 T; 5 other:
Query Match 2.1%; Score 82; DB 22; Length 613;
Best Local Similarity 48.4%; Pred. No. 5,9e-15;
Matches 257; Conservative 0; Mismatches 271; Indels 3; Gaps 1;
Oy 1614 TGTAGGTTAATTGATGCTCAGCTGTGTGTCAGTTCAGTGTGGGCTGAAGGTGAGC 1673
Db 37 TCGAGGATGATTTGGGCGGATGTGAGAGGGGACGGCTCATCTGTGAGGGGCGGCACT 96
Oy 1674 GCTGAACATCTCGGCTATGTTTGGCATTCTTCATGCACACATGCAATTCACACTTC 1733
Db 97 GACGTCACGGCTGGCTGTGATCTGTGAGAACGCTCTTCTTTCACATGAGGATCAGCAAGA 156
Oy 1734 TCCCTCGTTTGTACCGGGGTGTGCGGGCCGCTATTACAGTGGATGTTCCAGTGTGTG 1793
Db 157 CCCGATTTTGTACCAAGGCGCTGGGGGCTTATTTCTCAGCCATGTCACCTGGGTTCTG 216
Oy 1794 GTTAGTGAAGGCGGCAAGTGTGCGGAGCAGCTATTTGACCAAGCTACTTGATTTCGA 1853
Db 217 GCTGATGAGAGTGTGTCAGAGCGTTACTGGAATAATTGATAGACACATGCTACAGGCCA 276
Oy 1854 TCCGCTGTTGAAGAAAGCTCGGCAATGCGCACAGCTGTGAATCAGCCCTCCGTCGTG 1913
Db 277 TGCTGCTTTCCAGAACTCAAGTAAAGGCCACAGCCACATGCGCAGATATATATGACATA 336
Oy 1914 GCTTGCTATGCAATCTCGAAAAAAGCGGCACACCATCAGATGCTGTGCGCCCTGGCGAA 1973
Db 337 TTTGAACAGTCT--ACCTGATCTGATTAAGAAAGGCTCAGCCGTGGGCTTCTTACTGT 393
Oy 1974 AGGCTACACGTCGTCGGGAATTTCTGGAATATGCGGCGCCCTTCGAGATCTCATGTC 2033
Db 394 TGAATTACATGTTGGCCAGATTTCCATGCAACCGGTCTCCCTTAGCAGATGTGACACT 453
Oy 2034 CAGAGCGGTAATTTTGGCTGGGTATGAGAGCGAGATCTGATTAATTTACTGCGCTTGA 2093
Db 454 AAGGCGATGTCACCGGATTAAGACTGTCTCAGAGACCTGTGATGATCTTGCCATCTCTCA 513
Oy 2094 TATCGCTGATTAATGCGGATTTGTTATGTCGTGCGCCCAATTTCTGACGC 2144
Db 514 CCTGGCACAGTTCAAGGCAATGCTTGGGGAGCTGCTTCANTATATAGAGC 564

RESULT 15
AAH11917
ID AAH11917 standard; cDNA: 640 BP.
XX
AC AAH11917;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:4440.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
OS Aspergillus oryzae;
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
DR
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX
PS Claim 88; Page 1897; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organization of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
XX
SQ Sequence 640 BP; 166 A; 125 C; 183 G; 166 T; 0 other;
Query Match 2.0%; Score 78.4; DB 21; Length 640;
Best Local Similarity 53.4%; Pred. No. 8.6e-14;
Matches 191; Conservative 0; Mismatches 161; Indels 6; Gaps 1;
OY 1073 CACAAAGTGGGACGATCGGTTTGGATGCGCACCTGTCGNGGTACT-----GG 1126
DB 15 CAGAAACCGTACGTGGGATTGGCTTGTGATGACACATGTTCTGCTTCCACAG 74
OY 1127 ATAAAAACGGTGATCCATTGCTGTCTAGCCCGGAGGAGATGCAAAATATCATTTG 1186
DB 75 TTACCGCAGAACCTATATCTGTACGGGACCAATTTCGATTCGATCGCATGTCTATCC 134
OY 1187 TGTGATGATCTACCGGCGCACCGAACAGCAGAGCAATCAATGCCACTCACCCTCGG 1246
DB 135 TTTGGCTAGATCATCGCCCGTGGAGAGACGAAAGTTAATCTACCAATCATTAAT 194
OY 1247 TGCTGAACCTACGTGGGTAAATTTGGCTGAATGGAACACCGAAATTTCTGCGC 1306
DB 195 TGCTTGCTTACGTTGGTGGGAAATGTCCATTGAATGAGATTCCCAAGGTCCTGCGC 254
OY 1307 TGAAGAATAATATCCAGAGATCTACGAACGTCCGGACAATTTTTCATCTGGCGATT 1366
DB 255 TGAATAATCATATGCCCAAGATCTGTTCGACAGTCGCAAGTCTATGATCTGGCGATG 314
OY 1367 TTCTGACCTGGCGGCTACCGGTGATTTAGCGGTTCACTGATGCACTGTTACCTGTAA 1424
DB 315 CACTGACCATATTCGCAACTGCAATGAGAAAGAAATTTCTAGCGTTGTTGTAA 372

Search completed: March 21, 2003, 20:48:59
Job time : 910.004 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 09:44:37 ; Search time 143.991 Seconds
(without alignments)
8457.593 Million cell updates/sec

Title: US-09-802-208B-2

Perfect score:

Sequence: 1 atcgattgagcagtttgcct.....tcacgcgcgtcagccatcgat 3971

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```
IssuePatents_NA: *
1: /cgn2_6/prodata/1/lna/5A_COMB.seq.*
2: /cgn2_6/prodata/1/lna/5B_COMB.seq.*
3: /cgn2_6/prodata/1/lna/6A_COMB.seq.*
4: /cgn2_6/prodata/1/lna/6B_COMB.seq.*
5: /cgn2_6/prodata/1/lna/PCTUS_COMB.seq.*
6: /cgn2_6/prodata/1/lna/backfillseq.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	311	7.8	1401	4	US-09-134-001C-183	Sequence 183, App
2	69.6	1.8	696	4	US-09-134-001C-1560	Sequence 1560, App
3	68.8	1.7	744	4	US-09-385-028-22	Sequence 22, App
4	68.8	1.7	11604	4	US-09-385-028-13	Sequence 13, App
5	68.8	1.7	15079	4	US-09-385-028-1	Sequence 1, App
6	67	1.7	5133	4	US-09-453-702B-15	Sequence 15, App
7	57.2	1.4	5395	4	US-09-421-017B-383	Sequence 383, App
8	56	1.4	1155	3	US-08-793-035-2	Sequence 2, App
9	56	1.4	1185	3	US-08-793-035-1	Sequence 1, App
10	53.8	1.4	1512	1	US-08-504-808B-6	Sequence 6, App
11	52.2	1.3	1670	3	US-09-026-482B-1	Sequence 1, App
12	51.8	1.3	634	4	US-09-296-284-67	Sequence 67, App
13	51.8	1.3	684	4	US-09-296-284-66	Sequence 66, App
14	51.8	1.3	734	4	US-09-296-284-65	Sequence 65, App
15	51.8	1.3	784	4	US-09-296-284-63	Sequence 63, App
16	51.8	1.3	784	4	US-09-296-284-64	Sequence 64, App
17	51.8	1.3	834	4	US-09-296-284-62	Sequence 62, App
18	51.8	1.3	834	4	US-09-296-284-61	Sequence 61, App
19	51.8	1.3	934	4	US-09-296-284-60	Sequence 60, App
20	51.8	1.3	984	4	US-09-296-284-59	Sequence 59, App
21	51.8	1.3	1034	4	US-09-296-284-58	Sequence 58, App
22	51.8	1.3	1084	4	US-09-296-284-57	Sequence 57, App
23	51.8	1.3	1134	4	US-09-296-284-56	Sequence 56, App
24	51.8	1.3	1184	4	US-09-296-284-55	Sequence 55, App
25	51.8	1.3	1234	4	US-09-296-284-54	Sequence 54, App
26	51.8	1.3	1284	4	US-09-296-284-53	Sequence 53, App
27	51.8	1.3	1334	4	US-09-296-284-52	Sequence 52, App

28	51.8	1.3	1383	4	US-09-2366-284-51	Sequence 51, Appl
29	51.8	1.3	2700	4	US-09-2396-284-8	Sequence 8, Appl
30	50	1.3	1014	2	US-08-6731-190A-13	Sequence 13, Appl
31	49.4	1.2	750	4	US-09-134-001C-1988	Sequence 1988, Appl
32	45.6	1.1	763	4	US-09-221-017B-723	Sequence 723, Appl
33	45.6	1.1	5358	4	US-08-561-527-50	Sequence 50, Appl
34	42.4	1.1	31728	4	US-09-653-702B-64	Sequence 64, Appl
35	42.2	1.1	3058	4	US-09-221-017B-757	Sequence 757, Appl
36	41.8	1.1	741	3	US-09-238-481-1	Sequence 1, Appl
37	41.8	1.1	741	4	US-09-572-810A-1	Sequence 1, Appl
38	41.8	1.1	7218	1	US-08-232-463-14	Sequence 14, Appl
39	41	1.0	732	4	US-09-339-052-1	Sequence 1, Appl
40	40.4	1.0	825	4	US-09-134-001C-1594	Sequence 1594, Appl
41	39.4	1.0	19702	4	US-08-961-527-7	Sequence 7, Appl
42	39	1.0	1426	4	US-08-858-207A-78	Sequence 78, Appl
43	38.6	1.0	1086	1	US-08-464-400-1	Sequence 1, Appl
44	38.6	1.0	1086	3	US-08-875-273A-1	Sequence 1, Appl
45	38.6	1.0	1086	4	US-09-123-386-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-183

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

10 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATIN

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GIC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

;
PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 36/4
: SEQ ID NO 183

LENGTH: 1401

TYPE: DNA

ORGANISM: *Staphylococcus epidermidis*

FEATURE:

; NAME/KEY: unsure

LOCATION: (157)

OTHER INFORMATION
 HIS-09-134-001C-183

C
 C
 C
 F
 F
 C
 C
 F
 C
 F
 C

Query Match

Best Local Similarity 54.1%; Pred. No. 6.5e-90;
Matches 657; Conservative 0; Mismatches 555;
Indels 3; Gaps 1

QY	2589	TTGGGTTGGCCATCGCATGATGATGAGGATATATATGCGCATGCGAGTATATATAGCTGT	2648
		160	TTGGGTTGGCCATCGCATGATGATGAGGATATATATGCGCATGCGAGTATATATAGCTGT
Db	160	TTGGGTTGGCCATCGCATGATGATGAGGATATATATGCGCATGCGAGTATATATAGCTGT	219
QY	2649	GATGTTTGGAACTCGCATTTCTTATGCACTATATTAATCGTTAGGCTGCACACTCGG	2708
		220	GATGGAATGGAATACGCTTTCTCTCCGATATGATGATCTGGTCTTCTCCCTACT
Db	220	GATGGAATGGAATACGCTTTCTCTCCGATATGATGATCTGGTCTTCTCCCTACT	279
QY	2709	GAAAGCTCTTTGCCCTTACCGCTACAGGCGCTGGGCGCTGCCCTTCCGCCCTGGGTTCT	2768
		280	CAGGCAACCTCTTATTTATTTAGCGCTATGATGTTTTTGGCAGCCATTTCAGTTGGGGGTTC
Db	280	CAGGCAACCTCTTATTTATTTAGCGCTATGATGTTTTTGGCAGCCATTTCAGTTGGGGGTTC	339
QY	2769	GGGGTGTATGCGAAATCATCAAGCCGCAAAAAAACCATGCTGATTTGTCTTATG	2828
		340	GGTGTGTGGCGAAACCTTCGGTCTCGCGGAGATCATGCTGATTTGGCGTCCGGCTCG
Db	340	GGTGTGTGGCGAAACCTTCGGTCTCGCGGAGATCATGCTGATTTGGCGTCCGGCTCG	399
QY	2829	TGCGTATTCATCTCTGTTTCTGCGCTTTGGATTAGACAGGCAAACTATGGTTATC	2888
		400	ATTGCTCTTCATCTGTTTCTGATTTTTCGGCTCGGACACAGAAATATACATTATATG
Db	400	ATTGCTCTTCATCTGTTTCTGATTTTTCGGCTCGGACACAGAAATATACATTATATG	459


```

; STREET: The Jenifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELE: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-09-385-028-22

```

Query Match 1.7%; Score 68.8; DB 4; Length 744;

Best Local Similarity 47.4%; Pred. No. 1e-11; Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

```

QY 120 ATGAATACCTCCCTTAATGCAAGTTGCGAGTATCATCGGCTGCGTCAAGTATTGGC 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 744 ATGCCATCCGCACTCCAGGGAAAGTGGCGTCATCAGCGGCGGAGCTCGGCAATCGGC 685
QY 180 CTGCAATGTGCAGAAAAGCTGCTCGATGCAGAGCAAGAGTAGATTGATTGACGGGAA 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 684 GAGGCCAGCGCCCGGCGCTGCGCGCCGAGGCGCGCGCGGCGGCGGCGGCGGCGGCG 625
QY 240 GCGGCAAACTGCAGAAAGTTGTGCTGAGTT-----AGCGCAAAACGGTACGGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 624 GTGGAAGAACTGCGCGCTGGGTGACGAGTGAACCGCGCGGCGGAGGATCCATGTTC 565
QY 291 CTGCAACTGATCTCTTCAATATATGCAAGTGCATTAACATGCTGCGGAGCATTTATCGAA 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 564 CTGCAACTGACGTGCGCCACCGGCGGAGGGGTGACGCGCGCGTCCCTCCACCGTCGAG 505
QY 351 CTGGCGGGTGGGCTGATATTTTTCATGCCAATGCAAGCGCTTAATTTGGCGGCCAGTG 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 504 GGGCTGGGGGGGCTGACATCTCTGTCACAAAGCGCGGATCATGTGTCGGCCGGTG 445
QY 411 GCTGAAGTGATTCAGATGCTGCGATGCTGTGTTAAATCTGAATTAATATGCGGCGTTT 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 GAGGAGCGCGAGACCAACCGACTGACCCGGATGATGACCAACATCTCTGGGCGCTGATG 385
QY 471 GCGCTGTGTCGAGTCCCTGCGCATATGATGCGCAAGGTCGGGCGCATTAATTTT 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 TACATGACCCGGGCGGCTTCCCATCTGCT---GCGAGCAAGGGGACCGGTGGTGCAG 328
QY 531 ACCAGTTCCATCGGCGGCTGCTGCTGCTTATCTGGGAACCGATACACCGCGTACAAA 590
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 ATGTCTCTGATCGGCGGCGGAGTGAACGTCCGCAACGGCGCGTCTACAGAGCGACAGAA 268
QY 591 TTTCGGCTTCAGGATCTGTACACACTACCGCGCGCGAGTTTCTCAATATGCGCGT 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 TTCGGTGTGAACGCTTTCAGCAGACGCTGCCACAGAGAGTACACGAGCGCGGCGTGC 208

```

```

QY 651 GTGGGTGCGGTGCTGCGCAGACAGTAGTACTGCGCTGTTGATGAC 698
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 GTCTGTGATGAGCGCGGCGGACACCGGAGAGGAGTGGCGGCCAC 160

```

RESULT 4

US-09-385-028-13/c
Sequence 13, Application US/09385028
Patent No. 6232106

GENERAL INFORMATION:
APPLICANT: Susan E. Jensen

APPLICANT: Kwame A. Aidoo

APPLICANT: Ashish S. Parthkar

TITLE OF INVENTION: DNA sequence Encoding Enzymes of Clavulanic
Patent No. 6232106

NUMBER OF INVENTION: 25

CORRESPONDENCE ADDRESS:
ADDRESSER: JACOBSON, PRICE, HOLMAN & STERN, PLLC

STREET: The Jenifer Building, 400 Seventh Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462

FILING DATE: 29-JAN-1997

ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price

REGISTRATION NUMBER: 24,514

REFERENCE/DOCKET NUMBER: 1418/P57452US2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666

TELEFAX: (202) 39305350

IDEA UR

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 11604 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-385-028-13

Query Match 1.7%; Score 68.8; DB 4; Length 11604;

Best Local Similarity 47.4%; Pred. No. 6.9e-11; Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

```

QY 120 ATGAATACCTCCCTTAATGCAAGTTGCGAGTATCATCGGCTGCGTCAAGTATTGGC 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11333 ATGCCATCCGCACTCCAGGGAAAGTGGCGTCATCAGGCGCGGAGAGTGCATCGGC 11274
QY 180 CTGCAATGTGCAGAAAAGCTGCTCGATGCGAGAGCAAGAGTAGATTGATGACGGGAA 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11273 GAGGCCAGCGCCCGGCGCTGCGCGGAGGGGCGCGCGCTGCGCATTCGCGCGCGCG 11214
QY 240 GCGGCAAACTGCAGAAAGTTGTGCTGAGTT-----AGCGCAAAACGGTACGGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11213 GTGGAAGAACTGCGCGCTGGGTGACGAGTGAACCGCGCGGCGGAGAGTCCATGTTC 11154
QY 291 CTGCAACTGATCTCTTCAATATATGCAAGTGCATTAACATGCTGCGGAGCATTTATCGAA 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11153 CTGCAACTGACGTGCGCCACCGGCGGAGGGGTGACGCGCGCGTCCCTCCACCGTCGAG 11094

```

QY	351	CTGGGGGGTGGGGCTGGATATTTTTCATGCAATGCAAGGGCTTATATTTGGCCGACG	410
Db	11093	GCGCTGGGGCGGGCTGGACATCCTCGTCACAAACGCCGGGATCATGCTGCTCGGCCGGTG	11033
QY	411	GCTGAAGTGATCCAGATGTCTGGCATGCTGTGTTAAATCTGATATTAATGCGCGCTT	470
Db	11033	GAGGACGGCGGACACCACGACGACGCCGGAATGATCGAACCAATTCCTCTGGGCCCTGAG	10974
QY	471	CGCTTGTCCGGTGCAGTCTCTCGCGATATGATTTGGCAGAGTGGGGCATATTAATTTT	530
Db	10973	TACATGACCCCGGGGGGGCTTCCCATCTGCT---GCGAGGAAGGACAGCTGTGGCG	10917
QY	531	ACCAATTTCATTGGCGGGCGTCTTCCGGTTATCTGGGAACGATCTACACCCGTCGAA	590
Db	10916	ATGTCTCTGATCGCGGGCGGGGTGAACGTGCCAACGGCGGCGTCTTACACGGCCACGAAG	10857
QY	591	TTTGCCGTTTACAGGCAATTCGTACACACTACCCGCGCCGAGTTTCTCAATATGCGCGCT	650
Db	10856	TTTCGTTGTGAACGCTTACGAGACGCGTCCGCCAAGGAGTCAACGAGCGGGGGTGGG	10797
QY	651	GTGGGTGGGGTGCCTGCCAGAACCATATGACATGGCCGCTTGATGAC	698
Db	10796	GTCGTCGTCATGAGACCGGGACACACCGACACAGGAGAGCTCGCGCGCCAC	10749

RESULT 5
 US-09-385-028-1/c
 : Sequence 1, Application US/09385028
 : Patent No. 6232106
 :
 : GENERAL INFORMATION:
 : APPLICANT: Susan E. Jensen
 : APPLICANT: Kwamena A Aidoo
 : APPLICANT: Ashish S. Parachar
 : TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
 : Patent No. 6232106
 :
 : TITLE OF INVENTION: Acid Biotransformation
 :
 : NUMBER OF SEQUENCES: 25
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
 : STREET: The Jennifer Building, 400 Seventh Street, N.W.
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: U.S.A.
 :
 : ZIP: 20004
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30 (Epo)
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/385,028
 :
 : FILING DATE:
 : CLASSIFICATION:
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/790,462
 : FILING DATE: 29-JAN-1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: D. Douglas Price
 : REGISTRATION NUMBER: 24,514
 : REFERENCE/DOCKET NUMBER: 1418/P57452052
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 638-6666
 : TELEFAX: (202) 39305350
 :
 : TELEFAX: RCA 248593 IDEM UR
 :
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 15079 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 :
 : MOLECULE TYPE: DNA (genomic)
 : HYPOTHEetical: NO

```

; ORIGINAL SOURCE:
; ORGANISM: Streptomyces clavuligerus
US-09-385-028-1

Query Match 1.7% | Score 68.8 | DB 4 | length 15079;
Best local similarity 47.4% | Pred No. 8.3e-11;
Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2

```

QY	120	ATTAATACCTCCOCTTAATAGGCAAACTTTCAGCTATACCTGCGGCGTGGCGCAAGTATTGGC	179
Db	13365	ATGGCATCCGCACTCCAGAGGAAATCCGCTCATTCAGGCGGCACACTCGGCAATCGGC	133066
QY	180	CTGCATATGCAAAAACGCTGCTCGATCGAGGAGCAAAAGTATGATTGATTAGCCGGAA	239
Db	13305	GAGGCCACGACCGCCGCCCTGGCCGCCGAGGGCGCCGCGTGGCCATCGCCCGCGCGG	132466
QY	240	GCGGCAAACTGCGACAAAGATTGTCCCTAGTT-----AGCGAAAACGCGTACCGG	290
Db	13245	GTCGGAAGCTGCGCGGCCCTGGGTGACGACGTGACCGCGCGCGGCGGAAGTCCATGTG	13186
QY	291	CTGCACACTCGATCTCTTCAATATATGCAAGTCGATTAACATATGCTGGCGACATTATCGAA	350
Db	13185	CTCGAACTCGACGTGCGCGACCGGACGAGGGGTGAGACGCGCGCTCCACACCTCGAG	13126
QY	351	CTGGCGGGTGGGCTGGATATTTTTCATGCCAATGCAAGGCGCTTATATTGGCGGCCCATG	410
Db	13125	GCGGCGGGCGCGCTCGACATCCTCTGCAACAGCGCGGATCATGCTCGCGCGCGG	13066
QY	411	GCTGAAGTGTATCCAGATGTCTGGGATGCTGTGTTAAATCTGAATATTAATCGCGCGTTT	470
Db	13065	GAGGACGCCGACACACGACCTGAGACCCGAGATGACACCAATCTCTGGGCTGTATG	13006
QY	471	CGCTGTGTCGCGTGCAGTCTGCGCGCATATGATTGGCAGAGGTGCGGCGATTAATTTT	530
Db	13005	TACATGACCCCGGCGCGCTTCCCATCTGCT---CGGAGCAAGGCGACGCTGGTGGAG	12949
QY	531	ACCACTTCATCGCGGGCGTGTCTCGGTATCTGGGAACGATCTTACACGCGCTCCAAA	590
Db	12948	ATGCTCTCGATCGCGGGCGGTGAACGTCCGCAACGCGGCGCTTACACGCGCCAGAG	12889
QY	591	TTTGCGGTTCAAGGCAATGTGTAACAACATCCGCGCGCGGTTTTCATATATGGCGGCT	650
Db	12888	TTGCGTGTGAACGCGCTTACGACGACGCTCGCCACGAGGTTACCGAGCGCGGGTGGG	12829
QY	651	GTCGCTGCGGTCCTGCGACGACCACTAGTACTGCCCTCTGATGAC	698
Db	12828	GTCGCTGCTATCGACCGCGGACCAACGACAGGAGCTGCGCGGCGAC	12781
RESULT 6			
US-09-453-702B-15			
: Sequence 15, Application US/09453702B			
: Patent No. 6165723			
GENERAL INFORMATION:			
APPLICANT: Blattner, Frederick R.			
Burland, Valerie			
Perna, Nicole T.			
Plunkett, Guy			
Welch, Rod			
TITLE OF INVENTION: No. 6165723el Sequences of E. coli O157			
NUMBER OF SEQUENCES: 265			
CORRESPONDENCE ADDRESSES:			
ADDRESSEE: Quarles & Brady			
STREET: 1 South Plinkney Street			
CITY: Madison			
STATE: WI			
COUNTRY: US			
ZIP: 53701-2113			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Word Perfect 8.0			

```

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/453,702B
  FILING DATE: 03-DEC-1999
  CLASSIFICATION: <UNKNOWN>
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 60/110,955
    FILING DATE: 04-DEC-1998
  ATTORNEY/AGENT INFORMATION:
    NAME: Seay, Nicholas J.
    REGISTRATION NUMBER: 27386
  REFERENCE/DOCKET NUMBER: 960296,95017
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (608) 251-5000
    TELEFAX: (608) 251-9166
  INFORMATION FOR SEQ ID NO: 15:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 6133
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-453-702B-15

```

```

Query Match      -1.7%; Score 67; DB 4; Length 6133;
Best Local Similarity 47.7%; Pred. No. 1.7e-10;
Matches 236; Conservative 0; Mismatches 250; Indels 9; Gaps 1;

```

```

QY 119 TATGAACTACCTCCCTTAATGCAAGTTCGACCTATCATCTGGCGCTGCGCTAGATTGG 178
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 644 TTTGAGGAGGAACGTAATGATAGCTATTAATTAATTAATTAATTAATTAATTAATTAAT 703
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 179 CCTGAATGTGCAAAACGCTGCTGATGACGAGCAAGAGTAGTATTGATGACCGGGA 238
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 704 GGAAGGTATTGCGACAGAGCTTGGAATGACAGGTGCAAAAGTTTACTGGAGCACGCGAG 763
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 239 AGGCGACAAACTGCACAAAGATTGTCGCTGAGTT-----AGGCGAAAACGCGTAGCG 289
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 764 AGTTGAGCGTATCGAAGCATTTGCAACCGAATCTGCCGCGAGAGGAATTGCTAAAGC 823
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 290 GGTGAACCTCGATCTCTTCAATATCAGCAAGTTCGATACATGCTGGCGGACATTATGCA 349
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 824 GCGGAGATTGATGTCACAGACGACAGTCATGCGCGGATTCGTCGCAAGACGCGTGA 883
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 350 ACTGGCGGTGGGCTGATGATTTTTCATGCGCAATGCGAGGCGCTTATATTGGGCGCCAGT 409
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 884 TAGCTGGGGGCGAGTGTGATGTTCTTATCATATATGCGGCGTTATGCGCCTTACCGCT 943
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 410 GCGTGAAGGTGATCCAGATGTCGTGGATCGTGTGTTAAATCTGAATATTAATGCGCGCTT 469
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 944 TGCACAGGAAACAAAGATGATGAGGCGCTCAGTATGACGTGATATCAAAAGGTGTACT 1003
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 470 TCGCTGTGCTGCGTCAAGTCCTGCCCATATGATTTGCGGACAGAGTGGCGGATATAATTTT 529
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1004 GTGGGGAATTTGGCGCTGACTTCCGCTGATGGAACACACAGGCTTCGCGTCAAGATATCAA 1063
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 530 TACCACTTCATCGGCGGGGCGGTCCGCTTATCTGGGAACGAGATACAGCGCGTCAA 589
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1064 TCTTGTTCTATTGGTGCCCTTCTGTTGTGTGCCACAGCGCAGTCTATTGTGCATCAAA 1123
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 590 ATTTGCCGTTGAGC 604
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1124 ATTTGACAGTAGGCG 1138
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 7
US-09-221-017B-383/C
; Sequence 383, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120

```

```

CORRESPONDENCE ADDRESS:
  ADDRESSEE: MORRISON & FOERSTER
  STREET: 755 PAGE MILL ROAD
  CITY: Palo Alto
  STATE: CA
  COUNTRY: USA
  ZIP: 94304-1018
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: Windows
  SOFTWARE: FastSeq for Windows Version 2.0b
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/221,017B
    FILING DATE: 23-DEC-1998
  CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP1182
      FILING DATE: 31-DEC-1997
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP1546
      FILING DATE: 30-JAN-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP2911
      FILING DATE: 09-APR-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/AU98/01023
      FILING DATE: 10-DEC-1998
  ATTORNEY/AGENT INFORMATION:
    NAME: Montoy, Gladys H
    REGISTRATION NUMBER: 32,430
    REFERENCE/DOCKET NUMBER: 27340-20021,00
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 650-813-5600
      TELEFAX: 650-494-0792
    TELEX: 706141
  INFORMATION FOR SEQ ID NO: 383:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 5395 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: circular
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: UNKNOWN
    ORIGINAL SOURCE:
      ORGANISM: PORPHYROMONAS GINGIVALIS
      FEATURE:
        NAME/KEY: misc.feature
        LOCATION: 1...5395
US-09-221-017B-383

```

```

Query Match      1.4%; Score 57.2; DB 4; Length 5395;
Best Local Similarity 47.1%; Pred. No. 2.3e-07;
Matches 176; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 320 ACCTGATTAACATGCTGCGCGCATATATGCAACTGGCGGCGGTGGATATTTTCATGC 379
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1370 ACCACATACAGTGtGGAGACAGATCAAGCTCAGTTCGTCGATGATATTTTGGTCAA 1311
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 380 CAATGACAGGCGCTTATATTTGCGGCGCCAGTGCCTGAAGTGTATCCAGATGCTGGAGTCG 439
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1310 CAATGCGGCTATTACCGCGGATGAGCGCTTATGATGGTATGACCGAGCAACATATGGGATGC 1251
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 440 TGTGTTAATCTGAATATTAATGCGCGCTTTCGCTGTGCTGCGACAGTCTCCCGCATAT 499
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1250 CGTGATCAACGTTAATCTGMACTCAGCGCTTCAACATGATCCATGCTCCATCAT 1191
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 500 GATTGGCGAGAGTGGCGGATATATTTTTCAGATTCCATCGGCGGCGTTCGCGGT 559
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1190 GATGGACAGCGCACCGGAGTATCATATATGCTTGTGTGTAGGTGTATCGAGCAA 1131
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 560 TATCTGGGAACGATCTAACCGCGTCCAAATTTGCGGTTGAGGCAATTCGTACACACTAC 619
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db 1130 TGAGGACATCCAACTACCTCTTAAGCCGATATGCGTTGGCCAAAGACTGT 1071
QY 620 CCGCCCGGAGTTTCTCAATATGCGTGTGCGTGTGCGGACAGACTGT 679
Db 1070 CGCCAAAGAGCTGGCGTACAGTGTGCGTGTGCGCAATGCAATGCGGCTTTATCAT 1011
QY 680 CACTGCCCTGCTTG 693
Db 1010 CACCGATATGACTGT 997

RESULT 8
US-08-793-035-2
; Sequence 2, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Dianne
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketocyl ACP Reductase Genes from
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,035
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9414622.2
; FILING DATE: 20-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB95/01678
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713.787.1400
; TELEFAX: 713.787.1440
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-793-035-2

Query Match 1.4%; Score 56; DB 3; Length 1155;
Best Local Similarity 43.1%; Pred. No. 1.9e-07;
Matches 269; Conservative 0; Mismatches 355; Indels 0; Gaps 0;

QY 109 CTGTGCCCTTATGAATCTCCCTTAANGCAAAGTTGCGATATCATGCGCGTGGT 168
Db 209 CTGTCCGAAGAGTGAGTCTCCGCTGTGTGTGACTGTGCTGCAAGAGGATTGGTA 268
QY 169 CAGCTATTTGGCTGCAATGTCAGAAAAAGCTGCTGCATGCAAGAGCAAGAGTATGA 228
Db 269 AACCTATTTGCTTCTTCTGGGCAAGAGTGTGCAAGGCTTGTGTAAGTATGATAGT 328

QY 229 TTGACCGGGAAGGCGACAAACCTGCACAAGATTGCGTGAATTAGCGAAAGCGGTACG 288
Db 329 CAGCAAGAGAGGCTGAGGAAGTTTCTAAACAGATTGAAGCATATGAGGCCAGGCTATTA 388
QY 289 CGCTGCACTGCATCTCTTAATATATCAGCAAGTCCATTAACATGCTGGCGGACATATATG 348
Db 389 CTTTGGGGGTGATGTCCTCCAAAGAGGCTGATGTGAAGCCATGATGAAAAAGCGCTATTG 448
QY 349 AACTGGCGGTGGCTGATATTTTCATGCAATGAGGCGCTTATATGCGGCCAG 408
Db 449 ATGATAGGGAAGACCATTTGATGTGTGTCACAAATGCAAGCAATCCTGCGGATACCTTGT 508
QY 409 TGCGTGAAGGTGATCCAGATGTCTGGATCTGTAAATTCATATATATATGCGCGT 468
Db 509 TGATACGAATGAAGAAGTCCCAATGGATGAAGTATTTGAATCTCATCTGAGAGTCT 568
QY 469 TTGCTGTGTCCGTGCAAGTCTGCGCATATGATTTGGCGAGAGTGGCGGATATATTT 528
Db 569 TTCTGTGTACCCAGCGACAAAGATCATGTGAAGAGAAAGGAAGATCATCA 628
QY 529 TTACCAAGTTCATGCGCGCGCTGCTCCGTTATCTGGAACCGATCTACCCGCTCCA 588
Db 629 ACATTTGCTGAGTGTGTTGCTCATTTGTGTAATTTGGCCAAAGCAATACGCTGCTGTA 688
QY 589 AATTGGCGTTCAAGCATTCGTACACACTACCCGCGCAGGTTTCTCAATATATGCGTGC 648
Db 689 AACCTGCTTATTTGGGTTCTCCAAAGACTGCCCGCAGAGAGGCTGCGAGCATATATA 748
QY 649 GTGTGGTGGGTGCTGCTGCCAGCAAGTACTGCTGCTGTTGATGACTGGCCAAAG 708
Db 749 ATGTCAATGTGTGTTGGCTGCTCATTTGATGTGACATGACTGCCAAGCTTGGAGAG 808
QY 709 CCAAAATGGAAGAGCCCTGCA 732
Db 809 ACATGGAAGAAATCTTGGA 832

RESULT 9
US-08-793-035-1
; Sequence 1, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Dianne
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketocyl ACP Reductase Genes from
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,035
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9414622.2
; FILING DATE: 20-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB95/01678
; FILING DATE: 17-JUL-1995

OY 558 GTTATCTGGAGACGATCTACACCGCTCCAAATTTGCCCT 598
DB 739 CGTCGGGAATTTGGCCCTATATACGCGACCAAGAGGCGGT 779

RESULT 11

US-09-026-482B-1
Sequence 1, Application US/09026482B
Patent No. 6143538
GENERAL INFORMATION:
APPLICANT: REISER, STEVEN E.
APPLICANT: SOMMERVILLE, CHRIS
TITLE OF INVENTION: ACYL-COA REDUCTASE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL A. GOTTLIEB, AGCTT
ADDRESSEE: DEPARTMENT OF ENERGY
ADDRESSEE: GC-62 (FORSTL) MS-6F-067
STREET: 1000 INDEPENDENCE AVE. S.W.
CITY: WASHINGTON, D.C.
ZIP: 20585
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORDPERFECT 6.22
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,482B
FILING DATE: 02/19/98
ATTORNEY/AGENT INFORMATION:
NAME: ALMAN, JOY
REGISTRATION NUMBER: 40486
REFERENCE/DOCKET NUMBER: S-87814
TELECOMMUNICATION INFORMATION:
TELEPHONE: 630-252-2179
TELEFAX: 630-252-2779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1670 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-482B-1

Query Match 1.3%; Score 52.2; DB 3; Length 1670;

Best Local Similarity 47.5%; Pred. No. 4.1e-06;

Matches 233; Conservative 0; Mismatches 243; Indels 15; Gaps 2;

OY 133 TTATGCAAGTTGCAAGTATCACTGCGCTGCTCAGGATATGCTGCAATGTGCA 192
DB 395 TAAAGGTAAAGTGGCTTTGATCTACGTGCTCATCTAGTGAATGGTTGACGATTCGAA 454
OY 193 AAAGCTGCTGATGACGAGCAAGGATGATGATTGACCGGGAAGGCGCAAACTGC 252
DB 455 AAAGATTTGCTGGCGGAGGTCTCATATTTGTTGGCCGAACCAACCAAACTGCG 514
OY 253 ACAAGATTGCTGCTG-----AGTTAGCGAAACGCGTACGCGCTGCACACTGCATC 303
DB 515 AAGAAGTGAAGAGCTGCAATTGACGACGAAGGGGACAGGCGCTTATTTTCTTGTCGAC 574
OY 304 TCTTCATATATCAGCAGTGCATACATGCTGGCGGACATTTATGAACTGGCGGCTGGCC 363
DB 575 TGACTGACATGAATGCGATGACACGATTATTCACAACAATATGCGCAGTGTGCATCATG 634
OY 364 TGGATATTTTTCATGCAATGCAAGCGCTTATATTTGGCGGCGCATGTGAA-----G 417
DB 635 TCGATTTTCGATCAATAATATGACAGGCGCTTCGATTCGCGCGCTACACAGATCGTTTG 694
OY 418 GTGATCAGATGCTGCGGATCGTGTGTTAAATCTGAATATTAATGCGCGCTTTCGCTGTG 477
DB 695 ATGCTTCATGATTTTGAACGACCATGCAAGTGAATTTACTTTGGTGGCGGTACGTTTGA 754

OY 478 TCCGTGCAATCTCTCCGATATGATTCGCGACAGAGGTGCGGATATATTTTACCACTT 537
DB 755 TGTAAATTTTACTCCACATATGATTAAGCCTAAATAAAGCGACATATATATGAGCT 814
OY 538 CCATCGGCGGCGCTTCGCTTCCGTTATCTGGGAACCGATCTACACGCGGCCCAATTTGCCG 597
DB 815 CTATTTGCTGATTTGGCCAAATGCGACCCGTTTCTGCTTATGTGCGCTTAAACGCTCGC 874

OY 598 TTCAGGCAATTC 608
DB 875 TGGATGCGCTTC 885

RESULT 12

US-09-296-284-67
Sequence 67, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 67
LENGTH: 634
TYPE: DNA
ORGANISM: Glucobacter suboxydans
US-09-296-284-67

Query Match 1.3%; Score 51.8; DB 4; Length 634;

Best Local Similarity 50.6%; Pred. No. 2.8e-06;

Matches 125; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

OY 343 TTATGCAATGCGGCTGGATATTTTTCATGCCAATGACGCGCTTATATTTGGCG 402
DB 12 TGAACGACGCGTGGGCTGATTTGCTGCAACAATATGCGGCTCAGTTCCCTC 71
OY 403 GCCCAGTGGCTGAAGTGATGATGATGCTGCTGCTTAAATCTGAATTAATG 462
DB 72 GCCCTGCTGGTGAACAGACCGTGAAGAGTGGACACCGTGCAGGCCATTAACCTGCGTG 131
OY 463 CGGCGTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 522
DB 132 CGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 191
OY 523 TAAATTTTACCATGTTTCATGCGGCGCTGCTTCCGTTATCTGGAACCGATTCACCG 582
DB 192 TCATCAACATGATGCTCCAGCGACGCTGTCGCGCTGATGACCATGCTTACCTGCG 251
OY 583 CGTCAA 589
DB 252 CATCAA 258

RESULT 13

US-09-296-284-66
Sequence 66, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
;
US-09-296-284-66

```

Query Match	1.38;	Score 51.8;	DB 4;	Length 684;
Best Local Similarity	50.6%;	Pred. No. 3e-06;		
Matches 125; Conservative	0;	Mismatches 122;	Indels 0;	Gaps 0;

OY	343	TTATGAACTGGCGGTGGAGTATTTTTCATGCCAATGCAGGCGTTATTTGGCG	402
Db	62	TGAACGACGCTGGGGTGGGTATTTGTGTCAACAATCTGGCCGACTTCCCTC	121
OY	403	GCCCCAGTGGTAGAAGTATCCAGATGTCGTGGAGTCGTGTAAATCATTAATAATG	462
Db	122	GCCCCTCTGGTGGAAACAGACCCTGCAGAGATGGGAGCACCGATGCAGGCCATTAACTGGCGT	181
OY	463	GGGGTTTTCCGCTGTCCTCGTCGATGTCGCGCATATATGATTCGGCAGAGTGGCGGATA	522
Db	182	CGCCATGGCTTCTCCCGCTGTCTTTCGCTCCGGCATGATTAAACGAACGCTGGAAAA	241
OY	523	TAAATTTTACAGTTTCATCCGCGGCGTGTCTCCGGTATCTGGGNAACGATCTAACCG	582
Db	242	TCATCAACATCAGTTCCACGAGGCACGCTGTGCGCCTGATTGCATGAGTGGCTTAAGCTG	301
OY	583	CGTCCA 589	
Db	302	CATCCA 308	

RESULT 14
US-09-296-284-65

; Sequence 65, Application US/09296284A

```

: Patent No. 6204040
: GENERAL INFORMATION:
: APPLICANT: Choi, Eun-Sung
: APPLICANT: Rhee, Sang-Ki
: APPLICANT: Lee, Eun-Hae
: TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
: TITLE OF INVENTION: and Methods of Use Thereof
: FILE REFERENCE: 1533. 0870000
: CURRENT APPLICATION NUMBER: US/09/296,284A
: CURRENT FILING DATE: 1995-04-22
: NUMBER OF SEQ ID NOS: 87
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 65
: LENGTH: 734
: TYPE: DNA
: ORGANISM: Gluconobacter suboxydans
: US-09-296-284-65

```

Query Match	1.38;	Score 51.8;	DB 4;	Length 734;
-------------	-------	-------------	-------	-------------

Best Local Similarity 50.6%; Pred. No. 3.1e-06;
Matches 125; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Db	112	TGAACGACGCTGGGGTGGGTGGATATTTGTCGACCAATAGCTGGGCTCAGTTCCCTC	1717
Qy	343	TTATTCGAACATGGCCGGGTGGCTGGATATTTTTCATGCCCAATGACAGGGCTTATATTTGGC	407
Db	112	TTATTCGAACATGGCCGGGTGGCTGGATATTTTTCATGCCCAATGACAGGGCTTATATTTGGC	407
Qy	403	GCCACATGCGTGAAGGTGATCCAGATGTTGGGATCGTGTTAAATCTGATATAATG	463
Db	172	GCCCTCTGTGGAAACAGACCGTCGAGGAGTGGGACACCGCTACAGGCATTAACTGGGTG	231
Qy	463	CGGCGTTTCGCTGTGTCCGTGCTCAGTCTCTGCCGCAATATGATTTGCCCAAGTGGGCGATA	522
Db	232	CGCCATGCGCTTCGCGCCGCTGTTCTTGGCTCCGGGCATGATGAACCCAGCGTGGGAAAA	291
Qy	523	TAAATTTTAAACAGTTCCATCGCGGGGCGTCGTTCCGGTTATCTGGGAACCATCTACACG	583
Db	292	TCATCAACATAGTTTCCAGGCGACGCTCTGTCGGCGGTATGACCAAGTGCCTTACGTCG	351

QY	583	CGTCCA	589
Db	352	CATCCA	358

RESULT 15

; Sequence 63, Application US/09296284A

Patent No. 6204040
 GENERAL INFORMATION:
 APPLICANT: Choi, Eun-Sung
 APPLICANT: Rhee, Sang-Ki
 APPLICANT: Lee, Eun-Hae
 TITLE OF INVENTION: Glucanobacter Suboxydans Sorbitol Dehydrogenase, Genes
 TITLE OF INVENTION: and Methods of Use Thereof
 FILE REFERENCE: 1533.087000
 CURRENT APPLICATION NUMBER: US/09/296,284A
 CURRENT FILING DATE: 1999-04-22
 NUMBER OF SEQ ID NOS: 87
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 63
 LENGTH: 784
 TYPE: DNA
 ORGANISM: Glucanobacter suboxydans
 OS-09-296-284-63

Query Match 1.38; Score 51.8; DB 4; Length 784;

Best Local Similarity 50.6%; Pred. No. 3.3e-06;
Matches 125; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy	343	TTATCGAACATGCGGGGCGGTGATATTTTTCATGCCAATGCAAGCGCTTATATTGGC	402
Db	162	TGAACGAGCGTGGGGGGGTGGATTTGTGTCACAACTGCTGGGGTCAAGTTCCCTC	221
Qy	403	GCCCACTGCGTGAAGTGATCCAGATCTTGGGATCGTGTGTTAAATCTGAATTTAAATG	462
Db	222	GCCCTCGTTGGGAACAACCGCTCCAGAGAGTGGGACACCGCTTAACTCGCTG	281
Qy	463	CGGGTTTCGCTGTGTCGCGTCAATCGTGGCGCATATGATTTGGCCAGGCGGGGCGATA	522
Db	282	CGCCATGCGTCTTCGCCGATCTTTCGCTCCGGGCAATGATTTGAACGCAAGCGTGGAAAA	341
Qy	523	TAAATTTTACCAATTCATCGCGGGCGTGTCTTCGGTTATCTGGGAACCAATCTACACG	582
Db	342	TCATCAACATCATGTTCCACAGCCAGCTGTGCGCGGATTTGACCATGATGTACTACGTG	401
Qy	583	CGTCCAA	589
Db	402	CATCCAA	408

Search completed: March 21, 2003, 11:12:44
Job time : 186.991 secs

Search completed: March 21, 2003, 11:12:44
Job time : 186.991 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:05:17 ; Search time 296.853 Seconds
(without alignments)
10373.549 Million cell updates/sec

Title: US-09-802-208B-2

Perfect score: 3971

Sequence: 1 atcgattgagagtgctgtc.....tcacgcgcgtacatcatgat 3971

Scoring table:

Gapop 10.0, Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCF_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCFUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3971	100.0	3971	US-09-802-208B-2	Sequence 2, Appli
2	101	2.5	417	US-09-960-352-11317	Sequence 11317, A
3	71.2	1.8	1335	US-09-738-626-127	Sequence 127, App
4	67	1.7	6133	US-10-114-170-15	Sequence 15, Appl
5	64.6	1.6	789	US-09-815-242-6500	Sequence 6500, Ap
6	64.6	1.6	29729	US-09-070-927A-238	Sequence 238, App
7	63.4	1.6	468	US-09-974-300-2287	Sequence 2287, Ap
8	63.2	1.6	744	US-09-974-300-6558	Sequence 6558, Ap
9	58	1.5	807	US-09-712-363-99	Sequence 99, Appl
10	56.4	1.4	8320	US-09-813-453A-71	Sequence 71, Appl
11	55.6	1.4	6948	US-09-070-927A-4	Sequence 4, Appli
12	54.4	1.4	537	US-09-974-300-6537	Sequence 6537, Ap
13	50.8	1.3	1872	US-09-885-303A-15	Sequence 15, Appl
14	50.6	1.3	1714	US-09-885-303A-13	Sequence 13, Appl
15	50.2	1.3	680	US-09-974-300-2280	Sequence 2280, Ap
16	48.4	1.2	11427	US-09-070-927A-165	Sequence 165, App
17	48.8	1.2	3153	US-09-922-501-13	Sequence 13, Appl
18	48.2	1.2	735	US-09-815-242-6029	Sequence 6029, Ap
19	47.2	1.2	744	US-09-815-242-7795	Sequence 7795, Ap

20	45.8	1.2	756	10	US-09-974-300-2400	Sequence 2400, Ap
21	45	1.1	765	9	US-09-978-758-1	Sequence 1, Appli
22	44.4	1.1	774	9	US-09-891-6641-55	Sequence 55, Appl
23	44.4	1.1	1043	10	US-09-823-901-1	Sequence 1, Appli
24	43.4	1.1	741	10	US-09-815-242-8705	Sequence 8705, Ap
25	43.4	1.1	741	10	US-09-815-242-9002	Sequence 9002, Ap
26	42.4	1.1	31728	9	US-10-114-170-64	Sequence 64, Appl
27	42.2	1.1	672	10	US-09-974-300-6548	Sequence 6548, Ap
28	41.6	1.0	765	10	US-09-974-300-2256	Sequence 2256, Ap
29	41.4	1.0	14286	10	US-09-070-927A-162	Sequence 162, App
30	41	1.0	732	10	US-09-815-242-9262	Sequence 9262, Ap
31	41	1.0	732	10	US-09-815-242-9483	Sequence 9483, Ap
32	40.8	1.0	714	10	US-09-823-901-3	Sequence 3, Appli
33	40.8	1.0	717	10	US-09-974-300-2252	Sequence 2252, Ap
34	40.6	1.0	418	10	US-09-880-107-529	Sequence 529, App
35	40.4	1.0	708	9	US-09-738-626-2684	Sequence 2684, Ap
36	40.4	1.0	833	10	US-09-770-445-686	Sequence 686, App
37	40.2	1.0	599	9	US-09-978-758-11	Sequence 11, Appl
38	40.2	1.0	738	10	US-09-815-242-4406	Sequence 4406, Ap
39	40.2	1.0	741	10	US-09-815-242-8025	Sequence 8025, Ap
40	40.2	1.0	1905	10	US-09-922-488-3	Sequence 3, Appli
41	40	1.0	807	9	US-09-738-626-1473	Sequence 1473, Ap
42	40	1.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
43	39.8	1.0	762	10	US-09-974-300-2259	Sequence 2259, Ap
44	39.8	1.0	816	10	US-09-815-242-9289	Sequence 9289, Ap
45	39.8	1.0	816	10	US-09-815-242-9581	Sequence 9581, Ap

ALIGNMENTS

RESULT 1

US-09-802-208B-2

Sequence 2, Application US/09802208B

Publication No. US20030041352A1

GENERAL INFORMATION:

APPLICANT: Parriott, Wayne

APPLICANT: Lafayette, Peter

APPLICANT: Kane, Patrick

TITLE OF INVENTION: Antibiot. or Ribitol As Positive Selectable Markers

FILE REFERENCE: UGA 85SR

CURRENT APPLICATION NUMBER: US/09/802, 208B

CURRENT FILING DATE: 2001-03-08

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 3971

TYPE: DNA

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: misc_feature

LOCATION: (859)..(2463)

OTHER INFORMATION: ribitol dehydrogenase coding region

FEATURE:

NAME/KEY: misc_feature

LOCATION: (2565)..(3839)

OTHER INFORMATION: ribitol kinase coding region

NAME/KEY: misc_feature

LOCATION: (2565)..(3839)

OTHER INFORMATION: ribitol transporter coding region

US-09-802-208B-2

Query Match

Best Local Similarity 100.0%, Score 3971; DB 9; Length 3971;

Matches 3971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCGATTGACGAGTTGCTTCACACGCGCTAAATCCCGTTCAGTGCAGTCAACAA 60

|||||

DB 1 ATCGATTGACGAGTTGCTTCACACGCGCTAAATCCCGTTCAGTGCAGTCAACAA 60

|||||

OY 61 CAGAGCTATTAATTGCGCCCTGTAAAGATTATATGATGATCACTGTGCCCTCTA 120

|||||

Db 61 CAGAGACTAATAATTCGCCCTGGTAAAGATTATATGATGAATCACTGTGTGCTCTA 120
Qy 121 TGAATACCTCCCTTAATGCGAAAGTTGACAGTATCACTGGCGTGCCTGAGTATTGGC 180
Db 121 TGAATACCTCCCTTAATGCGAAAGTTGACAGTATCACTGGCGTGCCTGAGTATTGGC 180
Qy 181 TGCATATGCGAAAGAGCTGCTGATGTCAGAGCAAGGTAATGATTGATGACCGGAG 240
Db 181 TGCATATGCGAAAGAGCTGCTGATGTCAGAGCAAGGTAATGATTGATGACCGGAG 240
Qy 241 GCGACAACAGCGACAAGATTGTGCTGATGTTAGCGGAAAGCGGTACGGCTGCACATCG 300
Db 241 GCGACAACAGCGACAAGATTGTGCTGATGTTAGCGGAAAGCGGTACGGCTGCACATCG 300
Qy 301 ATCTCTTCATTAATACGCAAGTGAATACATGCTGGCGGACATTATGCACTGCGCGGTG 360
Db 301 ATCTCTTCATTAATACGCAAGTGAATACATGCTGGCGGACATTATGCACTGCGCGGTG 360
Qy 361 GGGTGGATATTTTCATGCGCAATGCGAGCGCTTATATTTGGCGGCCAGTGGTAGGTG 420
Db 361 GGGTGGATATTTTCATGCGCAATGCGAGCGCTTATATTTGGCGGCCAGTGGTAGGTG 420
Qy 421 ATCCAGATGCTGGATCGTGTGTTAAATCTGAATATTAATGCGGCGTTTCGGTGTGTC 480
Db 421 ATCCAGATGCTGGATCGTGTGTTAAATCTGAATATTAATGCGGCGTTTCGGTGTGTC 480
Qy 481 GTGCACTGCTGCGCATATGATGCGAGAGTGGCGGATATTAATTTTACCAGTTTCA 540
Db 481 GTGCACTGCTGCGCATATGATGCGAGAGTGGCGGATATTAATTTTACCAGTTTCA 540
Qy 541 TCGCGGGCGTGTTCGGGTATCTGGGAAACCGATCTACACCGGCTTCCAAATTTGCCGTC 600
Db 541 TCGCGGGCGTGTTCGGGTATCTGGGAAACCGATCTACACCGGCTTCCAAATTTGCCGTC 600
Qy 601 AGGCAATTCGTAACACATACCGCGCGGAGTCTCAATATGCGTGTGCTGGTGGTGG 660
Db 601 AGGCAATTCGTAACACATACCGCGCGGAGTCTCAATATGCGTGTGCTGGTGGTGG 660
Qy 661 TCGTGCAGAGCAAGTACTGCTGCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 661 TCGTGCAGAGCAAGTACTGCTGCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTG 720
Qy 721 AAGCCCTGGCAATGAGTACTGATGCAACGATTAAGTGGGGAATGATATTGTTTA 780
Db 721 AAGCCCTGGCAATGAGTACTGATGCAACGATTAAGTGGGGAATGATATTGTTTA 780
Qy 781 TGGTGACCCGCTGAAATATGTCACCGTGGAGATTTAGTATGCTGCTGCTGCTGCTG 840
Db 781 TGGTGACCCGCTGAAATATGTCACCGTGGAGATTTAGTATGCTGCTGCTGCTGCTG 840
Qy 841 ATCTGTAAGGCGCATATGACATTAACCAAAACGTTATTGGTGTGATGTTGGATCA 900
Db 841 ATCTGTAAGGCGCATATGACATTAACCAAAACGTTATTGGTGTGATGTTGGATCA 900
Qy 901 GGCAGTGTCCGCGCGGATTTTGTATCTCAACGATCTGTGATCCATGTCACAGAA 960
Db 901 GGCAGTGTCCGCGCGGATTTTGTATCTCAACGATCTGTGATCCATGTCACAGAA 960
Qy 961 AAAATACGACTACGCGCGCGAGCGGAAAGCGGTGGAACAGTCCAGCGAGATGTG 1020
Db 961 AAAATACGACTACGCGCGCGAGCGGAAAGCGGTGGAACAGTCCAGCGAGATGTG 1020
Qy 1021 CAGGCGGTCTGTTCAATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1021 CAGGCGGTCTGTTCAATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Qy 1081 GTGGCAGGCAATGCTGTTGATGCACTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 GTGGCAGGCAATGCTGTTGATGCACTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Qy 1141 CCATTCCTGTACCGCGGAGAGATGCAAAACCAATATCATTTGTTGATGATGATCAC 1200
Db 1141 CCATTCCTGTACCGCGGAGAGATGCAAAACCAATATCATTTGTTGATGATGATCAC 1200

Qy 1201 CCGCGCACCGAAGCAGAGCAATCAATGCCACTCCACATCCGGTCTGAATACATC 1260
Db 1201 CCGCGCACCGAAGCAGAGCAATCAATGCCACTCCACATCCGGTCTGAATACATC 1260
Qy 1261 GGTGTGTAATTTTCCCTGTAATGGAACACCGAAATTTCTGTGGCTGGAAGAAATATG 1320
Db 1261 GGTGTGTAATTTTCCCTGTAATGGAACACCGAAATTTCTGTGGCTGGAAGAAATATG 1320
Qy 1321 CCAAGATCTACGAAAGTCCGAGCAATTTTGTGATCTGGCGGATTTCTGACCTGGCG 1380
Db 1321 CCAAGATCTACGAAAGTCCGAGCAATTTTGTGATCTGGCGGATTTCTGACCTGGCG 1380
Qy 1381 GCTACCGGTATTTAGCGCTTCAGTATGCACTGTTACTGTAAATGAGAGTGGTGCA 1440
Db 1381 GCTACCGGTATTTAGCGCTTCAGTATGCACTGTTACTGTAAATGAGAGTGGTGCA 1440
Qy 1441 CATGAAATATGCTGGGATCCAGATTTATTTCCGACCATTTGGCCTTGAGAGTTAGCG 1500
Db 1441 CATGAAATATGCTGGGATCCAGATTTATTTCCGACCATTTGGCCTTGAGAGTTAGCG 1500
Qy 1501 GAAGATTTTATTCGATTTGCTATCATATTTGTTTCTCCGGAACACCTTGGAATATGT 1560
Db 1501 GAAGATTTTATTCGATTTGCTATCATATTTGTTTCTCCGGAACACCTTGGAATATGT 1560
Qy 1561 TTAACAGCACAAGCGCGGAGAGATGGATTAATCTCCCGGCAACCTGTGCTGTAGGG 1620
Db 1561 TTAACAGCACAAGCGCGGAGAGATGGATTAATCTCCCGGCAACCTGTGCTGTAGGG 1620
Qy 1621 TTAATGATGCTACGCTGGTGGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 TTAATGATGCTACGCTGGTGGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Qy 1681 AATCTGCGTATGTTTTCGCGACTTCTTCAATGCAACATGAGCATCTACACTTCTCCCTG 1740
Db 1681 AATCTGCGTATGTTTTCGCGACTTCTTCAATGCAACATGAGCATCTACACTTCTCCCTG 1740
Qy 1741 TTTGTACCGGCTGTGCGGCGGCTGATTAACAGTGGATGATGATGATGATGATGATG 1800
Db 1741 TTTGTACCGGCTGTGCGGCGGCTGATTAACAGTGGATGATGATGATGATGATGATG 1800
Qy 1801 GAAGCGGCGCAAGATGCTGCGGAGAGCAATGATGATGATGATGATGATGATGATGAT 1860
Db 1801 GAAGCGGCGCAAGATGCTGCGGAGAGCAATGATGATGATGATGATGATGATGATGAT 1860
Qy 1861 GTTGAAGAGCTGCGCAATGAGCAAACTGTGATGATGATGATGATGATGATGATGATG 1920
Db 1861 GTTGAAGAGCTGCGCAATGAGCAAACTGTGATGATGATGATGATGATGATGATGATG 1920
Qy 1921 GATCGAATCTCGCAAAAGAGGCGCAACATGATGATGATGATGATGATGATGATGATG 1980
Db 1921 GATCGAATCTCGCAAAAGAGGCGCAACATGATGATGATGATGATGATGATGATGATG 1980
Qy 1981 CAGTGTGCGGCAATTTCTCGGAATTCGCGCCCTTCGCAATCTCATATGCGAGAGCG 2040
Db 1981 CAGTGTGCGGCAATTTCTCGGAATTCGCGCCCTTCGCAATCTCATATGCGAGAGCG 2040
Qy 2041 GATATTTGTGCGCTGGGATGAGCGAGATCTGATATTTTACTGCTGCTGTATATGCT 2100
Db 2041 GATATTTGTGCGCTGGGATGAGCGAGATCTGATATTTTACTGCTGCTGTATATGCT 2100
Qy 2101 GGAATTATGCGGAATTTGTTATGCTGCGCAAAATTCACAGCTCAAAAGCGAGAGGA 2160
Db 2101 GGAATTATGCGGAATTTGTTATGCTGCGCAAAATTCACAGCTCAAAAGCGAGAGGA 2160
Qy 2161 GTAGTGAATTAATATATGTTATGAGCGGAGTGGCGGAGATCCAGTGTAGAGCA 2220
Db 2161 GTAGTGAATTAATATATGTTATGAGCGGAGTGGCGGAGATCCAGTGTAGAGCA 2220
Qy 2221 ATTCTGCGAGATACCTGCGGATTTCCGCTATTAACAGCAAGTGTGCGAAGCGGTTT 2280
Db 2221 ATTCTGCGAGATACCTGCGGATTTCCGCTATTAACAGCAAGTGTGCGAAGCGGTTT 2280

OY	2281	TTAGGCTGGGCAATTCGTGGGTGCTGCGCGGAAATATTGCAACCTTCGTTGGGGAAGC	2340
Db	2281	TTAGGCTGGGCAATTCGTGGGTGCTGCGCGGAAATATTGCAACCTTCGTTGGGGAAGC	2340
OY	2341	ATGCACAATTCCACCACATGCGATGATAAATATTAATTCGCGAAGAACGCTATCAATCTCTT	2400
Db	2341	ATGCACAATATTCACCCATGCGATGATAAATATTAATTCGCGAAGAACGCTATCAATCTCTT	2400
OY	2401	CATCATCTGTCGATATGAGCGCTATAAGCAGTTGCACACATCTGCAAAATTAACAGAGAC	2460
Db	2401	CATCATCTGTCGATATGAGCGCTATAAGCAGTTGCACACATCTGCAAAATTAACAGAGAC	2460
OY	2461	TAAATTAACGACGCGCGAGCGGTTCCTTAGGCAATACATAATACGACACACGCGGTA	2520
Db	2461	TAAATTAACGACGCGCGAGCGGTTCCTTAGGCAATACATAATACGACACACGCGGTA	2520
OY	2521	ATATCCCGAGTGCATTCATCTGACACCTTAACAAAAGAGGTTTATATCTCACAAATATA	2580
Db	2521	ATATCCCGAGTGCATTCATCTGACACCTTAACAAAAGAGGTTTATATCTCACAAATATA	2580
OY	2581	AACAGTGGTTGGTTTGCCACTGCACTGATATATGGGATATATCCGCAATCCAGTATTTA	2640
Db	2581	AACAGTGGTTGGTTTGCCACTGCACTGATATATGGGATATATCCGCAATCCAGTATTTA	2640
OY	2641	TGACATGCTGATNGTTTGCAGACCTGCACTTCTTAATCCAGACTATATTAACGTTAGGCTTCA	2700
Db	2641	TGACATGCTGATNGTTTGCAGACCTGCACTTCTTAATCCAGACTATATTAACGTTAGGCTTCA	2700
OY	2701	CACCTGGGGAAGCCCTTTTGGCTTAACGCTCTACGGGCTGCGGCTCCCTTCGCGCT	2760
Db	2701	CACCTGGGGAAGCCCTTTTGGCTTAACGCTCTACGGGCTGCGGCTCCCTTCGCGCT	2760
OY	2761	GGGTTTCTGGGGTAGTAGCGGAATCATCAGCCGCGAAAAACATGCTGATTTGGTTTG	2820
Db	2761	GGGTTTCTGGGGTAGTAGCGGAATCATCAGCCGCGAAAAACATGCTGATTTGGTTTG	2820
OY	2821	TCCATATGTCGCTATTCACATGTTCTGTCTTCTGCTCTTGATAGGACAGCAACTATG	2880
Db	2821	TCCATATGTCGCTATTCACATGTTCTGTCTTCTGCTCTTGATAGGACAGCAACTATG	2880
OY	2881	GCTTAATCTGCGTGTTCCTTACGGGATTCGAGGTTTAGCGATACGGTATTTCTACACAT	2940
Db	2881	GCTTAATCTGCGTGTTCCTTACGGGATTCGAGGTTTAGCGATACGGTATTTCTACACAT	2940
OY	2941	TTATTCGTCGTTATTTATTCATTAACGTGCGTAGGAAAAATCCAGTTCTGCTCGGGTGGT	3000
Db	2941	TTATTCGTCGTTATTTATTCATTAACGTGCGTAGGAAAAATCCAGTTCTGCTCGGGTGGT	3000
OY	3001	ACTGGGCGGATATATTCAGTAGGAGATGGGTGCTGCGGCAATTATTCGCCAGTTTACGA	3060
Db	3001	ACTGGGCGGATATATTCAGTAGGAGATGGGTGCTGCGGCAATTATTCGCCAGTTTACGA	3060
OY	3061	TACGATTAATGGGGTAATGGGAACCTTAATNGCTGGCACTGGCGTTCGCTTCTGTGGCG	3120
Db	3061	TACGATTAATGGGGTAATGGGAACCTTAATNGCTGGCACTGGCGTTCGCTTCTGTGGCG	3120
OY	3121	GTCGATATGGCATTGATCTCCCTTCGCTCAGTAAATTAACGCGCTGGAATATATGCATATTTAA	3180
Db	3121	GTCGATATGGCATTGATCTCCCTTCGCTCAGTAAATTAACGCGCTGGAATATATGCATATTTAA	3180
OY	3181	CTCCCGCTGAGAAAGTTTGGCAGAAATTAAGTCGGGCGAGTAACTTACTTATATCCAAACGCA	3240
Db	3181	CTCCCGCTGAGAAAGTTTGGCAGAAATTAAGTCGGGCGAGTAACTTACTTATATCCAAACGCA	3240
OY	3241	ATATTTTTCCTCCAGTATTTGTCGGAATTAATAACCTTATCGTATTCGTTATTCGGTTTGGG	3300
Db	3241	ATATTTTTCCTCCAGTATTTGTCGGAATTAATAACCTTATCGTATTCGTTATTCGGTTTGGG	3300
OY	3301	TCATATTAAGCAATGATCTTTGTGGATGAACCTGGGATTCACACCTCTGATATGTTGGACG	3360
Db	3301	TCATATTAAGCAATGATCTTTGTGGATGAACCTGGGATTCACACCTCTGATATGTTGGACG	3360
OY	3361	TCGCGGCGGCAATTTTCTTACCACTAATTTCTCTAATAATTTTGGGGGATGTGGCG	3420

Db 3361 TCTGGCGCGCAATTTTCTTACCACTATTTTCTCTAATATTTTGGGGGATTTGGCAG 3420
 QY 3421 AAAAAATGGCGTGATGCGGTATATTCGCTGGTTGGTTCGCGGATGGCAGCATCAA 3480
 Db 3421 AAAAAATGGCGTGATGCGGTATATTCGCTGGTTGGTTCGCGGATGGCAGCATCAA 3480
 QY 3481 GTTTAGCGTTTACTACTAGCCGCAATACTTTGGTCACACTACTGATGGCAATGATTC 3540
 Db 3481 GTTTAGCGTTTACTACTAGCCGCAATACTTTGGTCACACTACTGATGGCAATGATTC 3540
 QY 3541 CGGCGATTGCTCTGGGAACCTTTGTGCTGCATTTGGCCGATGGCGCTGTCTTCCCG 3600
 Db 3541 CGGCGATTGCTCTGGGAACCTTTGTGCTGCATTTGGCCGATGGCGCTGTCTTCCCG 3600
 QY 3601 CACTGGAACCAAAACACAAAGGTGCTGCATCTCGGTTTACACCTCTGCGGGATATGT 3660
 Db 3601 CACTGGAACCAAAACACAAAGGTGCTGCATCTCGGTTTACACCTCTGCGGGATATGT 3660
 QY 3661 CTAACTTCTTGCGTCCGCAATTCGCCGTGGTTATTACGTGGTTAGCACTATCCGCTG 3720
 Db 3661 CTAACTTCTTGCGTCCGCAATTCGCCGTGGTTATTACGTGGTTAGCACTATCCGCTG 3720
 QY 3721 TGGTCATTGCGCTATACAGCATTTGATCTATTTGGCCTTGTGCTTGGGCATCTATTCGCG 3780
 Db 3721 TGGTCATTGCGCTATACAGCATTTGATCTATTTGGCCTTGTGCTTGGGCATCTATTCGCG 3780
 QY 3781 TTGAGCAGCCAGATTCAGTTCTCTGCGCAGTACTGAGAAAGCATTAATATTCCTGAA 3840
 Db 3781 TTGAGCAGCCAGATTCAGTTCTCTGCGCAGTACTGAGAAAGCATTAATATTCCTGAA 3840
 QY 3841 AAACGAAAGCATGACGACATCTCTCTCCCTCATGGGAGAGAGATGTTTACATCAGG 3900
 Db 3841 AAACGAAAGCATGACGACATCTCTCTCCCTCATGGGAGAGATGTTTACATCAGG 3900
 QY 3901 CAATAGTACTTTGTTATGAGATTAAGCTCTGACGGCTTATAGTTTACGGCGGT 3960
 Db 3901 CAATAGTACTTTGTTATGAGATTAAGCTCTGACGGCTTATAGTTTACGGCGGT 3960
 QY 3961 CAGCCATCGAT 3971
 Db 3961 CAGCCATCGAT 3971
 RESULT 2
 US-09-960-352-11317
 : Sequence 11317, Application US/09960352
 : Patent No. US20020137139A1
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Wesley C.
 : APPLICANT: Tao, Nengping
 : APPLICANT: Byatt, John C.
 : APPLICANT: Mathalagan, Nagapan
 : TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 : FILE REFERENCE: 16511.006/37-21(10298)C
 : CURRENT APPLICATION NUMBER: US/09/960.352
 : CURRENT FILING DATE: 2001-09-24
 : NUMBER OF SEQ. ID NOS: 15112
 : SEQ ID NO 11317
 : LENGTH: 417
 : TYPE: DNA
 : ORGANISM: Bos taurus
 : OTHER INFORMATION: Clone ID: 48-LIB34-080-Q1-EL-DB
 US-09-960-352-11317
 Query Match 12.5%, Score 101, DB 10, Length 417,
 Best Local Similarity 56.2%, Pred. No. 2, 7e-22;
 Matches 232; Conservative 0; Mismatches 175; Indels 6; Gaps 2;
 QY 1194 GGATACCGGCCCAACGAAACAGACAGACGATCAATGCCACTCACCATCCGGTGTGAA 1253
 2 GGATACCGGCCGCGTCACTAGGTCCACCGGATCAACGAGACAGACGATCAACGCGTCTGCA 61

D	155	CCCTCGGGGAAACCTCGTAGCTGTTTACGGTGTGTCGGGGGAGTCGACGATTCCTTG	214
Q	2767	CTGGGGTAGTAGCCGAATATCATACGCCGCAAAAAAACATGCTGATTTGGTCTAT	2822
D	215	CGGGCGACATTTGGACCTCTGGGGGTCCAGAAAGTAGATCTCGGGTCTTCATCT	274
Q	2827	GGTCCGATTTCCATAGTCTCTTTCTTGCTCTTGATAGAGCGCAACATAGGGTTAA	2886
D	275	GGATTCGCTTTGACCTGATCTCTCTACCGCTGGCACTGACACCGACATACTGGTTGA	334
Q	2887	TTCCTGCTGTTTTACGGGATTCGTTGTTAGCGTATCCGCTATTTCTCTACTATTATCG	2944
D	335	TCTTCCTTGGCTTAGGTCCTCGGGGGCTTTGGTTATCCATCTTCGCTACGGATTCGG	394
Q	2947	TGCTTTATTAATCATTAAGCTGCGTAGAGCAAAATTCACAGTTCTGCTGGGTTGGTACTGG	3006
D	395	TGTGGATCACCGCAACTGCTCACTACATAGCAATTTGGGTACCGGTGTGGTTGTTCTACG	454
Q	3007	CGGTATATTCAGTAGGGATCGGTGTGGTCGACGATTATATCCAGTTTTCAGATACGA	3066
D	455	TTTGCTTCTCTGCGAGGCTTCTTCACTTTGGGTGCGGTGGTGGCACATTTTCATAGCA	513
Q	3067	TTATGGGTGAATAGGAAACCTTATAGGTGACATGGCGTTCGCTTGGTGGCGGT-GTC	3122
D	514	---TACGTGACTTGGACCTTCTATGAAGCTTTGGGGTTTCCCTCGTGTGGTGGTCATC	570
Q	3126	ATTGCACTGATCTCTGCGCTCATGTTTAAACGCTT-----GACATATGATATATTA	3175
D	571	GGATCGCTCATCGCACTGCGAGTAGTAAGAACGTGCGGAGCCACCCACTGGTGGCC	630
Q	3180	ACTCCCGGTAGAAAGTTTGGAGATTAAGTCGGGGCAGTAACTTACTTATACCAACGC	3233
D	631	AACCCGACAGATGTAAGCAACACTTGGCCAGGGCTTCAAACTTTCGCGCATGTATGA	690
Q	3240	AATATTTTCTCTCCACTATTGTGCGCATATATTAATACCTTATCGTTATTCGGTTTGG	3295
D	691	CGTGCAGGTTTTGTACACTCAATCCGACCATCACTCCATTCGACCTATACCGATGGGT	750
Q	3300	GTCATTTATGCCATGATGTTTGTGGATGAACCTGGGATTACCACTCTCATATGTTTCAG	3355
D	751	GTTGTTCTCCATCATATTTTCACTGACGATCTAAGTGCAGCTAAGCTGATTCCTCATC	810
Q	3360	GTCGGGCGGCAATTTTCTTTTACCACTAATTTCTCAATATTTTGGGGATTTGGCA	3415
D	811	CTCACCACTGTAATTTACGAGATCAACACTGCCCTCAATCTTTCTTTCGGTAGCTTCGG	870
Q	3420	GAATAAAATGGCTGAGTGCSTGTTATTTCCGTGTTGTTGGTCTGGGATGCGACATCA	3475
D	871	GACCGCACAGTTGGGACCAACACTGTGTTCTGGGGCGGATCAATCGTGGCCAGCTACC	930
Q	3480	AGTTTACGTTTTACTACATGACCGCGCAATCTTGGTCA-----C	3518
D	931	CTCGCGTTGTTTACTTCTCATTCGAGATGTCGGCGTTCAAGCTGGCATGTGCCAAGTGC	990
Q	3519	AACACTGATGCGCATGATTTCCGGGAGTTGCTCTGGAACTTTTGTGCTCATTTTGG	3578
D	991	GTTTTCGGATACACATCGCAGCGCGGAGCACTTTGGTGTGTCCTCGCGGCTTGGG	1056
Q	3579	CCGATGGCGCGCTGTCTCCCGGCACTGGAACCAAAACAAAGGTGTGCAATCTCGGT	3638
D	1051	CCACTTTTCCGCAATCGCTGTCTCCCTTGTATCCCAACACCCCGGGCGACGATGTGCCA	1110
Q	3639	TACAACTCTCTCGGGGTATGTTACTCTTCGCGCTCCGGCAATTTGCGGTGTTATTA	3698
D	1111	TACAACTCGGGCTTGGTGGCGCTGTAGCTGTGGAGCCGCTCTGTGGTTCAGTCTTTCAC	1170
Q	3699	CCGTGTTTGAACATATCGGTGTGGTCAATTTGCCATATACAGATTTGATCTATTTGGCTTT	3758
D	1171	CCACTGATTTGGTCCAAACCGGATGATCTCGTGATCATGATGAGCCCTCTACCTGCTCTCGGT	1230
Q	3759	GTCCTTTGGCAGTATCATTTGCGCTTGTAGACGACCAAGATTTAGTTTTCGCGCAG	3810
D	1231	TGAGTACCTTTCACCTTGCAGCTGCGGCAACCAACAGATTTTGACAGGTGCGAG	1282

Db 80 AGATTGCCAATCCCTAGCTGAAGCAGACACAAATTTGTTTTTAATTAATCTATCTCTG 139
QY 244 ACAAACTGCAAGATTGTGCTGAGTT-----AGCGCAAAACGCGTACGCGCTGC 294
Db 140 AATCAGTTGAACAAGCTTGGAATTAATGCGCGCAGGAGATTGAAGCACTGCTATG 199
QY 295 AACTGATCTCTTCAATATCAGCACTGCATTAACATGCTGGCGGACATTAATGCACTGG 354
Db 200 TGTGTGATGTCTAGTGAAGAGCAAGTCAAGCGATGTCTGCTCAATTAAGAAGAAAG 259
QY 355 CGGCTGAGCTGATTTATTTTTCATGCGCAATGAGCGCTTATTTGGCGCCAGTGGCTG 414
Db 260 TCGGCTTATTTGATTTTATTTTACTATATATCGCGAATTATTAACGAATTCATGCTGG 319
QY 415 AAGGTATCCAGATGTCGTGGAGTCTGTAAATCTGAATTAATTAATGCGCGCTTTCGCT 474
Db 320 ATATGTGACAGAAAGATTCCTGCAAGTGAATGATGATGATGATGATGATGATGATGATG 379
QY 475 GTGTCCGTGATCTCTGCGCATATGATGCGCAGAGGTCGGCGGATTAATTTTACCA 534
Db 380 TGGCAAAAGGGGTATTCAGATATGATGAAAAAGGTCATGCGCAAAATTAATCAATATCT 439
QY 535 GTTCATCGCGGCGG 549
Db 440 GTTCATGATGATGAGTG 454

RESULT 6
US-09-070-927A-238
Sequence 238, Application US/09070927A
Patent No. US20020120116A1

GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 238:

SEQUENCE CHARACTERISTICS:

LENGTH: 29729 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 238:
US-09-070-927A-238

Query Match 1.6%; Score 64.6; DB 10; Length 29729;
Best Local Similarity 48.7%; Pred. No. 7.3e-09;

Matches 212; Conservative 0; Mismatches 214; Indels 9; Gaps 1;

QY 124 ATACTCCCTTAATGCGAAAGTTGACATTCATGCGCGCTGCTGAGTATGGCTGC 183

Db 15623 ATTTTCATTAATGAGAAAGTACGCTTAATACAGTCTGTTATGAGATTGGCTTTG 15682

QY 184 AATGTGCAAAAGCGCTCTGATGACAGACCAAGAGTATGATGATGATGATGATGATG 243

Db 15683 AGATTGGAAATCCCTAGCGAAAGACCAACAAATGTTTTTAATCACTTATCTCTG 15742

QY 244 ACAAACTGCAAGATTGTGCTGAGTT-----AGCGCAAAACGCGTACGCGCTGC 294

Db 15743 AATCAGTTGAACAAGCTTGGAATTAATGCGCGCAGGAGATTGAAGCAGCTGTTATG 15802

QY 295 AACTGATCTCTTCAATATCAGCAAGTCAATGATGCTGGCGGACATTAATGCAACTGG 354

Db 15803 TGTGTGATGTCTAGTGAAGAGCAAGTCAAGGATGTTGCTCAAGATTAAAGAAAG 15862

QY 355 CGGCTGAGCTGATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 414

Db 15863 TCGGCTCTATTTGATTTTATGCAATATATGCGGAATTAATTAACGAATTCATGCTGG 15922

QY 415 AAGGTATCCAGATGTCGTGGATCTGTAAATCTGAATTAATTAATGCGCGCTTTCGCT 474

Db 15923 ATATGTGACAGAAAGATTCCTGCAAGTGAATGATGATGATGATGATGATGATGATG 15982

QY 475 GTGTCCGTGATCTCTGCGCATATGATGCGCAGAGGTCGGCGGATTAATTTTACCA 534

Db 15983 TGGCAAAAGGGGTATTCAGATATGATGAAAAAGGTCATGCGCAAAATTAATCAATATCT 16042

QY 535 GTTCATCGCGGCGG 549

Db 16043 GTTCATGATGATGAGTG 16057

RESULT 7
US-09-974-300-2287
Sequence 2287, Application US/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods For Monitoring Multiple Gene

FILE REFERENCE: 10085.500-US

CURRENT APPLICATION NUMBER: US/09/974,300

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: 09/680,598

PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2287

LENGTH: 468

TYPE: DNA

ORGANISM: Bacillus licheniformis

US-09-974-300-2287

Query Match 1.6%; Score 63.4; DB 10; Length 468;
Best Local Similarity 46.8%; Pred. No. 6.9e-10;

Matches 199; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 174 ATTGGCCTGCAATGTGCAAAACGCTGCTGATGACAGACCAAGATGATGATGATGATG 233

Db 34 ATGCGCCTGACCTGCGGAAACGAGCAATGTCTGCTCAACTACGCGGGAATGAA 93

QY 234 CGGGAAGCGCAAACTGCAAGATTTGCTGAGTTAGGCAAAACGCGTACGCGCTG 293

Db 94 GCGAAGCAAGCAAGTCTAGACGAATCAAGCGCTGGCGCGATGGCTTTT 153
Oy 294 CAACGTGATCTCTCAATATATGACGAAGTGCATATGCTGGCGACATATGCACTG 353
Db 154 AAAGCGAGCTTCCATCGGATGAGTTCAGCGCATATGGAAGGCGGACGC 213
Oy 354 GCGGCGGCTGATATTTTTCATGCAATGACGCGCTATATTTGGCGCCAGTGGCT 413
Db 214 TTGGGACCGCTTACATCTCTTGCACCAATGCGGCGATTACTTAAGCAATCTTTCATG 273
Oy 414 GAAGTGATCAGATCTGCGATCGTGTAAATCTGAATTAATGCGCGCTTTCG 473
Db 274 AGAATGAAGAAGATGAATGGGACGACGTCATTAACATTAAGTTAAAGTGTTCAT 333
Oy 474 TGTGTCGTCGATCTCTGCGCATATGATTTGGCGACAGTGGCGCATATATTTTACC 533
Db 334 TGTTCAAAAGCTGTGCAAGACAGATGATGAACAAAGACGCGCATCAATATC 393
Oy 534 AGTTCATCGCGCGCTGCTCGTTATCTGGGAACCGATCAACCGCGTCAATTT 593
Db 394 ACCTCGGTGTAGCGCTCTGCTGTAACGCCGGGACGCACTATGTGCGCTTAATCA 453
Oy 594 GCGCT 598
Db 454 GCGCT 458

RESULT 8

US-09-974-300-6558
; Sequence 6558, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6558
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6558

Query Match 1.6%; Score 63.2; DB 10; Length 744;
Best Local Similarity 45.4%; Pred. No. 1.1e-09;
Matches 227; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Oy 267 GAGTTAGCGAAGCAAGCGTACGTCGACATCTCTCAATATATGCAAGTCAT 326
Db 151 GCGCTTGGCCAGAGGCGTTTGGCTATCAGGCTGACGTTCCCTCCAAAAGGAGTCGCA 210
Oy 327 AACATCTGCGGACATTTATGCACTGCGGCGTGGCTGATATTTTTCATGCAATGCA 386
Db 211 GCGATGATGAAGAGAGATGGCGGTTTCAATCAATGATATTTTATGAGCAATGCC 270
Oy 387 GCGGCTTATATTTGGCGCCAGTGGCTGAAGTGCATGATGTCGGATGCTGTGA 446
Db 271 GCGATTAACGCGCATATCTGTAATGAGATGAAGAAAGACGATTGGGATGCGCATTT 330
Oy 447 AATCTGAATTAATGCGGCTTTCGCTGTGTCGCGACATCTCCGCGCATATGATTCG 506
Db 331 GATACCAACTTAAAGAGGCTTCCATTTGCGCAAAAGCACTACGCGCAATATGATGAAA 390
Oy 507 CAGAGTGGCGGATATATTTTACAGTTCCATGCGGCGCTGCTTCCGTTATCTGG 566

Db 391 CAACGTGTCGCGAAGTATCATATGTCGTCTGTGCTTGGCGTAATGGGTAAAGCTGG 450
Oy 567 GAACGATCTACACCGCGCTCAATATTTGCGCTTACAGCATTCGTACACACTACCCGCG 626
Db 451 CAAGCGAATTAATGTTCCCGCAAGAGAGTGTATGCGTGTGATAGCTTTGGCGCG 510
Oy 627 CAGTGTTCATATATGCGCGTGTGGGTGGCTGCTGCGCAGACAGTACACTGCC 686
Db 511 GAATTTGCGAGGCGAGCATTTCTTTCATATGCGGTGCGCGCTTTATTAAGACAGAT 570
Oy 687 CTGCTGATGACTGGCCAAAAGCAAAATGGAAGAACCCCTGCAATGATGATCCGATG 746
Db 571 ATGACAGATGAGCTACCAAGCGAAGCAAGAACAGACAGTGTCTTAACAAATCCATTGG 630
Oy 747 CAACGATTAAGTGGCGGA 766
Db 631 AAGCTTGTGAACACAGGA 650

RESULT 9

US-09-712-363-99
; Sequence 99, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-99

Query Match 1.5%; Score 58; DB 9; Length 807;
Best Local Similarity 48.8%; Pred. No. 6.3e-08;
Matches 157; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Oy 378 GCGAATGCGAGCGCTTATATGCGCGCCAGTGGCTGAAGTGCATGCTGATGCTGGAT 437
Db 271 GCGAAGCGGCTACCGCGACATTTGCGCGCATGCTGCTGCTGCGCGGAAAG 330
Oy 438 CGTGTGTAATGTAATATATATGCGCGCTTTCGCTGTGCTGCGAGTCTGCGCAT 497
Db 331 AGCGAGTGAAGTGAATGCGCGGTGCGGTCAGACCTTACGTTAGTGGCGGTGCGCGCG 390
Oy 498 ATGATGCGCAGAGTGGCGGATATATTTTACAGTTCCATGCGGCGCTGCTCG 557

Db	351	ATGATCAGCGCCAAAGGCGCGGGCATCTTGATTTCTGTGTTGCGGGCGCGCAATTCACG	450
Oy	558	GTTATCTGGGAACCGATGTACACCGCGTCCAAATTTGCCGTTTACGGCATTCGTACACACT	617
Db	451	ATTCCCTCAACAACGGCACCTTACCGCGACCAAGGCTCTTGGAACACCTCTACAGGAATCT	510
Oy	618	ACCGCGCCGCAAGGTTTTCATATATGGGGTGCGTGCTGGGTGGGTCGCCAGGACCACTA	677
Db	511	CTGGCGGGGTGAGTACGCGCGCTCGGGGGTGACGTCAACGGGTGCTGCCCGCGCCGCGTT	570
Oy	678	GTCACTGCCCTGCTTGATGACT	699
Db	571	CGCACCGAGCTACCGGATGCTT	592

```

US-09-813-453A-71
US-09-813-453A-71
Sequence 71, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:-
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227, 860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667, 569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 71
LENGTH: 8320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: plasmid, PAN296
US-09-813-453A-71

```

	Query Match	Similarity	1.4%	Score 56.4	DB 9	Length 8320
	Best Local	Similarity	47.7%	Pred. No. 1.3e-06		
	Matches	165	Conservative	0	Mismatches 181	Indels 0
QY	347	CGAACTGGCCGGTGGCGCTGGATATATTTTTCATGCCAATGACGGCGCTTATATTTGGCGGCC	406			
Db	639	CGATCAGATGGCGCTCGATGTATGTAAGTATTAACAATGACAGGCTTCGTATATTTGAAC	698			
QY	407	AGTGGCTGAAGGTGATCCAGATGCTGGGATCGTGTGTGTTAAATTCGAATATAATCGCC	466			
Db	699	GTTTTTAACACTTCATATGGATGACATGAAACGATGTTTGATGTGAATGTCTTCGGCCT	758			
QY	467	GTTTTGGCTGTGTCGCTGCAGCTCCGCCGATTTGATTTGGCAGAGGTCCGGGATATTAAT	526			
Db	759	GATCCCTGCTACAAAAGCGGTCTTCGCCAAATGCTTGAGCAGAAAAAGGACATATTCAT	818			
QY	527	TTTTACGATTCATCGCGGGCGTGCCTCGTTACTTGGGAACCGATCTACACCGCGTC	586			
Db	819	CAATATCGCCTCTCAAGCGGGGAAATATGCCACACCGAAGTCTAGCGTATATTCGCCGAC	878			
QY	587	CAAAATTTGCCGTTCAAGGCAATTCGTACACACTAACCCGCCGCAAGTTTCTCAATATATGCGGT	646			
Db	879	CAAAATCATCCGCTGTTAGGTTACTCAACAGCCTTTGGCGGATGAGGACTTTCGGGAACCGCAT	938			
QY	647	GCGTGTGGGTGGCGTGCCTGCCAGACCAAGTAACTACCTGCCCTGCTT	692			
Db	939	TTATGTGACAAACATGCAACCCCGGGCCGCAATTCAGACAGGACTTTT	984			

RESULT 11
US-09-070-927A-4/c
; Sequence 4, Application US/09070927A

Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,555
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301)309-8504
TELEFAX: (301)309-8512
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6948 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-070-927A-4

Query Match	1.4%	Score 55.6	DB 10	Length 6948
Best Local Similarity	46.4%	Pred. No. 2.1e-06		
Matches 218	Conservative	0	MissMatches 249	Indels 3
				Gaps 1
Qy	141	AAAGTTGACAGCATACACCTGGCGCTCGATCAGGATATATGGCCCTCAATGTCCAAAAACGCTG	200	
Db	3137	AAAGTTATTTGTTATCATATGGGCGCTTCACATGGGATCGGTGAGGACACACCCCTTACCT	3078	
Qy	201	CTCGATGCAGCAGCAAAAGTAGTATTGATTGAACCGGGAAGCGACAACTGCACAAGATT	260	
Db	3077	GCCAAAGAAAGGAGCAAGTAGTACTCATTTGCAGACACGTCGACAAAGACGTTAAATTGCTATA	3018	
Qy	261	GTCCGTGAGTTAGGGGGAACCGCG---TACGGGCTGCACACTGATCTCTTCAATATATCAG	317	
Db	3017	AAAAAAGCACTTCGGAAGCAACGATTCTCGTGCACAAACAGACGATTAACGAAAAGAAAG	2958	
Qy	318	CAAGTCGATTAACATCTCGCGGACATTTATCGAACTGGCGGGGTGGGATATTTTTCAT	377	
Db	2957	GAAGTTACGCGTGTAATCAAGCTTACATATGAAAAATATATGACGAATTTATGTCTTTTT	2898	
Qy	378	GCCAAATGCAGCGGCTTATATTATTTGGCGGCCACAGTGGCTGAAGGTATCCAGATGTTGGGAT	437	
Db	2897	AACATGTCAGGAGGTATATGCCAACGGCGGCTCTTATTGAAGCACACAAAAGAGAAATGGCGT	2838	
Qy	438	CGTGTTTAAATCTGAATTTAAATATCGCGGCTTTCCTGTGTCCGTGCATCTCTGCCGCAT	497	
Db	2837	CAATGCTGATGATTTTAAATTAATTAATGAGGTGTTTAAATATGCGATTGGCGGCACTTCTCCAATT	2778	

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:18:35 ; Search time 6094.69 Seconds
(without alignments)
10552.182 Million cell updates/sec

Title: US-09-802-208B-2

Perfect score: 3971

Sequence: 1 atcagatgagcagttgtc.....tcacgcgcgcacatcatgcat 3971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthu:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199.4	5.0	692	13	BI260394
2	189.2	4.8	922	14	B0934625
3	185.8	4.7	926	14	B0644899
4	171.4	4.3	736	13	BI759992
5	169.8	4.3	905	14	B0440371
6	169.8	4.3	1069	14	BM921107

7	156	3.9	676	14	BM791109
8	153.8	3.9	535	17	AZ302589
9	150	3.8	655	17	AZ650403
10	150	3.8	712	13	BI557944
11	149.2	3.8	714	13	BI149014
12	148.4	3.7	503	12	BF440790
13	147.8	3.7	676	9	AI527876
14	144.4	3.6	670	13	BM646051
15	142.4	3.6	983	14	B0932942
16	140.6	3.5	699	12	BF783891
17	139	3.5	853	12	BF218299
18	137.6	3.5	796	12	BF678897
19	135.8	3.4	654	9	AI108240
20	135.4	3.4	666	9	AI388552
21	135	3.4	857	13	BI329415
22	134	3.4	479	9	AI196200
23	132.8	3.3	669	14	B0406350
24	131.4	3.3	658	13	BI629061
25	130.8	3.3	621	10	AM164681
26	123.2	3.1	602	9	AI789136
27	122.2	3.1	1079	13	BI819693
28	120	3.0	533	9	AL749866
29	117.8	3.0	1053	11	AK09249
30	117.4	3.0	730	9	AL529279
31	117.4	3.0	851	14	B0892439
32	116.2	2.9	637	13	BI565802
33	115.6	2.9	634	10	AM107094
34	114.4	2.9	536	13	BM036183
35	112.8	2.8	732	13	BI613997
36	111.8	2.8	615	14	BQ801483
37	111.2	2.8	604	13	BI588685
38	111.2	2.8	612	13	BI577185
39	109.6	2.8	386	14	BM855062
40	109.4	2.8	907	12	BC184645
41	107.6	2.7	1095	17	CNS06C90
42	105.6	2.7	531	10	BE682075
43	104.4	2.6	582	13	BI583725
44	104.2	2.6	486	14	BM856746
45	104.2	2.6	795	9	AU121672

ALIGNMENTS

RESULT 1
BI260394
LOCUS 602969456F1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5108988 5',
DEFINITION mRNA sequence.
ACCESSION BI260394
VERSION BI260394.1 GI:14818651
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 692)
NIH-MGC <http://mgc.mci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contract: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue/Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Place: LAM11263 row: p column: 13
High quality sequence stop: 692.
Location/Qualifiers

FEATURES
source

1. 692

Query Match	Best Local Similarity	Matches	Score	DB	Length
1086	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
25	60.9%;	361; Conservative	199.4;	13;	692;
1146	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
85	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1206	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
145	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1266	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
205	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1326	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
265	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1383	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
325	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1443	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
383	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1503	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
442	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1563	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
502	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
1623	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;
562	5.0%; 60.9%;	361; Conservative	199.4;	13;	692;

JOURNAL Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: DCTD/DP/Gazdar
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L10CM2598 row: k column: 14
 High quality sequence stop: 723.
 Location/Qualifiers
 1..922
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6422149"
 /clone_lib="NIH-MGC-18"
 /tissue_type="Large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pORF7; Site: 1: XhoI; Site: 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally clone
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCCAGAG(C). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH-MGC Library."
 BASE COUNT 220 a 205 c 275 g 218 t 4 others
 ORIGIN
 Query Match 41.8%; Score 189.2; DB 14; Length 922;
 Best Local Similarity 56.4%; Pred. No. 1.3e-44;
 Matches 433; Conservative 0; Mismatches 322; Indels 13; Gaps 4;
 Oy 879 TATTGGTGTGGTGTGGATTCAGCAGTGTCCGCCCGGAGTTTTCATCTCAACGATC 938
 Db 163 TGTGGGTGTGGACGTTGGACACAGCAGTGTCCGTGTGACGTCGTGTGGACACAGTGGGT 222
 Oy 939 TCTGCTATCCACATCGCCACGAAAAATATCATGACGTCACGCGCGCAGCGGAGCGCGTGA 998
 Db 223 CCGTGTGGCTTTTGGACAGACAGCCAAATTAAGATTTGGAGCCCGCATTCACACCATGA 282
 Oy 999 ACAGTCCAGCCGAGGATCTGGCAGCGCGGTCTGTCATGATTCGTAATGGCGTCACCT 1058
 Db 283 GCAGTCTCCGAGACATCTGGCGTGTGCTGTGTGTTCACAAAGAAAGT-----TGT 336
 Oy 1059 GGCACAGCTTGTGTGACAAAGTGTGGCAGGCATCGGTTTGTATGTCACCTGTTCTGT 1118
 Db 337 ACAAGGAGATGTATTAACCAAAATTCGAGAGCACTTGAGGTTTGTATGTCACAGTGTCTGT 396
 Oy 1119 GGTACTGATTAACACGATGATTCATTTGCTGTCTACGCCCGGAAGAGATGCAACCAAAA 1178
 Db 397 TGTTTTGGATTAAGCAGTTTACCCATTCACCAATTAACCAAGAGAGGAGATTCACATGAAA 456
 Oy 1179 TATCATTTGTGTGATGAGATCACCGCGCACCAACAAAGACAGAGCAATCATATGCATCA 1238
 Db 457 CGTATCATGATGTGCTGGACACATCGACAGCATGCTCAAGTTAACAGATCAATGAAACCAA 516
 Oy 1239 CCATCCGGTGTGAACTACGTGGTGTGTAATTTTGCCTGAATGAAACACCGAAAT 1298
 Db 517 GCACAGTGTCTCCAGTACGTGGGGGGGTGTGTCTGTGTAAGTACAGGCCCCCAAACT 576
 Oy 1299 TCTCGGCGGAAGAAATATGCGAGAGATCAAC---GAAGTGGCGGCAATTTTGTGA 1355
 Db 577 TCTGTGGCTGAAGAACTTGAGAGATTTGTGGGATTAAGGGGGGCAATTTTCTTGA 636
 Oy 1356 TCTGCGCATTTTTCAGACTGGCGGGCTACCGGTAATTTAGCGCGCTTCAAGTATGACATGT 1415
 Db 637 TCTCCCGGACCTTCTATGCTGTGAAGCAACAGGTCTCACAGCAGGTCTCTGTGCTCCT 696
 Oy 1416 TACCTGTAATGACAGTGGCTGGACATGAAATATCGTGGATTCAGATTATTTCCGAC 1475

```

Db      697 GGTGTGTAAGTGACATATTCAGCA--NAGAAAGCTGGGACGACAGATTTCGAAAT 753
Qy      1476 CATTCGCTTGCAGATTAAGCGATGATGATTTTATTCGATTCGATCATATTTGTTTC 1535
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      754 GATGTGTTTGAAGACATTTGTTGAGATTAATTCACAGANNATAGG-AAACAATGCTACC 812
Qy      1536 TCCCGGACACCTTGTGGAATGTTTAAACAGCAGCCGCGAGAGATGGATTAAT 1595
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      813 TCCTGGAGCTTCTCTGGNATATGGGCTCACACGAGACAGACAGACCTTGGCTTCTC 872
Qy      1596 CCCCCGACACCTGTGCTGTAGGTTAATGATGCTCAGCTGTGG 1643
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      873 CCGGGGATTTGGGTCCTTCACTCATTTGATGCCCATGCGGGGG 920

RESULT 3
B0644899
LOCUS      B0644899      926 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8493114 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300239
5', mRNA sequence.
ACCESSION  B0644899
VERSION    B0644899.1 GI:21769071
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 926)
            NIH-MGC http://mhc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M2514 row: c column: 24
            High quality sequence stop: 546.
            Location/Qualifiers
                location=926
FEATURES             source
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:6300239"
     /clone_1ib="NIH_MGC_100"
     /tissue_type="hepatocellular carcinoma, cell line"
     /lab_host="DH10B (phage-resistant)"
     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
     into EcoRI/XhoI sites using the following 5' adaptor:
     GGCACAG(G). Size-selected >500bp for average insert size
     1.8kb. Library constructed by Ling Hong in the laboratory
     of Gerald M. Rubin (University of California, Berkeley)
     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
     II RT (Life Technologies). Note: this is a NIH-MGC
     library."
BASE COUNT      222 a      213 c      267 g      223 t      1 others
ORIGIN
Query Match      4 7%; Score 185.8; DB 14; Length 926;
Best Local Similarity 60.2%; Pred. No. 1.4e-43;
Matches 345; Conservative 0; Mismatches 222; Indels 6; Gaps 2;
Qy      1106 CCTGTTCTGTGCTGACTGATATAAAGCGTGATCATTCGCTGCAGCCGGAAGAG 1165
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 CGGTCTCTGTGTTTGGATTAAGCAGTTTCACCATTAACAGTCACACAGAGAGGGG 60
Qy      1166 ATGCAAAAGCAATATCATTTGTGTGATGATCAGCCGCCACCGAACAAGACAGCGAA 1225
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      61 ATTCCCATGAAACGTCATCATGTGGCTGACCATGAGCAGTCAGTCAAGTTAACAGGA 120

```

```

Qy      1226 TCATGCCACATCCATCCGGTGTGTAACACTACGCTGGTAAATTTGCGCTGAATGG 1285
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 TCATATGACACAGACAGATGCTCTCCAGTACGTCCGGGGGGATGTCGTGTGAAATGC 180
Qy      1286 AAACACCGAAATTTCTGTGGCTGAAGAAATATGCCAGAT--CTACGAACGTGCCG 1342
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      181 AGCCCGGAAACTCTGTGGCTGAAGAGAACTTGAGAGAGATTTGCTGGGATTAAGCGCG 240
Qy      1343 GACAATTTTGCATTCGCGGATTTTCTGACCTGCGCGGCTACCGGATTAAGCGCGTT 1402
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      241 GACATTTCTTTGATCTCCGCGACTTCTTATCGTGAAGGCAACAGGTGCACAGCGGT 300
Qy      1403 CAGTATGACATGTACTCTGTAATGAGACGTGGCTGCACATGAATAATCCGTGATTCAG 1462
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      301 CTCTCTGCTCCCTGGTGTGTGTAAGTGCATATTCAGCA---GAGAAAGCTGGGACAGCA 357
Qy      1463 ATTATTTCCGACCATTTGGCTTGCAAGTTAGCGAGTAAGATTTTATTCGATTTGTC 1522
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      358 GTTGTGGAATAATGATTTGTTGGAAGACTTGTTCAGATTAATTAACGAAATAGGAA 417
Qy      1523 ATCATATTGTTTCTCCCGGACACCTTTGTGAATGTTTAAACAGCAACAGCCGCGAG 1582
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      418 ACCAAGTGTACTCCTCGGAGCTTCTTGAATGGGCTCACACGACGAGCAAGAG 477
Qy      1583 AGATGGATTAACCTCCCGGACACACTGTGCTGTAGGCTAATTTGCTCAGCGTGGT 1642
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      478 ACCTTGCCCTTCTCCCTGGATTTGGGATTCGCGACCTTCATCTATGATGCCATGACAGAG 537
Qy      1643 GCATCGTACGTTGGCGTGAAGGTGAGAGCGC 1675
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      538 GACTAGAGTGAATTTGGGCGAGATGTGAGAGGC 570

RESULT 4
B1759992
LOCUS      B1759992      736 bp      mRNA      linear      EST 25-SEP-2001
DEFINITION 60304443F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184755 5',
mRNA sequence.
ACCESSION  B1759992
VERSION    B1759992.1 GI:15751570
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 736)
            NIH-MGC http://mhc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM11461 row: e column: 12
            High quality sequence stop: 736.
            Location/Qualifiers
                location=736
FEATURES             source
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:5184755"
     /clone_1ib="NIH_MGC_116"
     /lab_host="DH10B"
     /note="Organ: pooled colon, kidney, stomach; Vector:
     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
     source anonymous pool of 3 colons, age 26 yo male, 49 yo
     female, 71 yo male colon; 46 yo male kidney, and pool of 2
     stomachs, 62 yo male and 70 yo female. Library is

```



```

ACCESSION   BM921107
VERSION      BM921107.1  GI:19371486
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 1069)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Email: c9apbs-remail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12786, row: k, column: 06
             High quality sequence stop: 726.
FEATURES
             Location/Qualifiers
             1..1069
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:5752445"
             /clone_1lb="NIH_MGC_115"
             /lab_host="DH10B"
             /note="Organ: pooled brain, lung, testis; Vector:
             pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
             source anonymous pool of 6 male brains, age range 23-27; 1
             male lung, age 27; and 1 male testis, age 69. Library is
             oligo-dT primed and directionally cloned (EcoRV site is
             destroyed upon cloning). Average insert size 1.8 Kb,
             insert size range 1-3 Kb. Library is normalized and
             enriched for full-length clones and was constructed by C.
             Gruber (invitrogen). Research Genetics tracking code
             021. Note: this is a NIH-MGC Library."
BASE COUNT   233 a      270 c      304 g      262 t
ORIGIN
Query Match      4.3%; Score 169.8; DB 14; Length 1069;
Best Local Similarity 58.7%; Pred. No. 9.8e-39;
Matches 335; Conservative 0; Mismatches 227; Indels 9; Gaps 2;
QY 879 TATTGCTGTGATGATGCGATCAGCAGTGTCCGCCGGGATTTTGATCTAACGGATC 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 227 TGTGGGTGTGAGCTGTGGAACAGCTGTCCGTGACCTGTGTGACAGAGTGGGT 286
QY 939 TCTGCTATCCATGCCACAGAAAATCAGACTACGGCGCGCAGCGGAAGCCGCTGA 998
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 287 CCTGTGGCTTTTGCAGACACCACTTAAGATTGGAGCCCCCACTTACACCATCA 346
QY 999 ACAGTCCAGCAGAGATCTGCGAGCGGTCTGTTCTATGTTCTAATGCCCTACTCT 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 347 GCAGTCTCCGAGACATCTGGCTGCGTGTGTGTGTCACAAAGAAGT-----TGT 400
QY 1059 GGCACAGCTTTTGCACAAAGTGTGGCAGGCATCGTTTGTATGCCACCTGTTCTGT 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 401 ACAAGGATGATTAAACCAAAATTCGAGACTTGGGTTGATGCCACGTTCTCTGT 460
QY 1119 GGTACTGATTAAGAAAGCGATGATGCTGTGACGCCGGAAGAGATGCAAGCAAA 1178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 461 TGTTTTGAATAGAGATTTCACCATTTACAGTCAACAGAAAGGATTTCCCATGAAA 520
QY 1179 TATCATTTGTGTGATGATACCCGCCACCAAGCAAGAGCGAATGATGCACTGA 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 521 CCTCATCATGTGCTGCTGACCATGACAGCTCAAGTTAAACAGAGATCAATGAGCAA 580
QY 1239 CCATCCGGTCTGAGCTAGTGGGTGTAAATTTTGGCTGAATGGAACACCGAAAT 1298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 581 GCACAGTGTCTTCAGTACGTGGGGGGGTGATGTTGTGAATATGACGCCCGAAACT 640

```

```

QY 1299 TCTCTGCTGAAGAAATATGCGACAGAT---CTACGAACGTGCCGACAAATTTTGA 1355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 641 TCTGTGGCTGAAGAGAACTTGAGAGATTTGCTGGGATTAAGCGCGACATTTCTTTGA 700
QY 1356 TCTGGCGATTTTCTGACCTGGCGGCTTACCGTGATTAAGCGCTTCAGTATGCACTGT 1415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 701 TCTCCCGACCTTCTTATGCTGGAAGCAACAGTGTCAACAGCAGGTCTCTGCTCCCT 760
QY 1416 TACCTGTAAATGACGTGCTGCGACATGAA 1446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 761 GGTGTGTAAGTGACATATTCAAGACAGAA 791
RESULT 7
BM791109
LOCUS      676 bp mRNA linear EST 05-MAR-2002
DEFINITION K-EST0071070 S21SNUS20 Homo sapiens cDNA clone S21SNUS20-19-B10 5',
            mRNA sequence.
ACCESSION  BM791109.1  GI:19139341
VERSION     BM791109
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 676)
AUTHORS      Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
             Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
             Kim,Y.S.
TITLE        21C Frontier Korean EST Project 2001
JOURNAL      Unpublished (2002)
COMMENT      Contact: Kim YS
             Genome Research Center
             Korea Research Institute of Bioscience & Biotechnology
             52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
             Tel: +82-42-860-4470
             Fax: +82-42-860-4409
             Email: yongsung@email.kribb.re.kr
             Plate: 19; row: B column: 10
             High quality sequence stop: 676.
FEATURES
             Location/Qualifiers
             1..676
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="S21SNUS20-19-B10"
             /clone_1lb="S21SNUS20"
             /sex="F"
             /tissue_type="Stomach"
             /cell_line="SNU-520"
             /lab_host="Top10F"
             /note="Organ: Stomach; Vector: pTZ18Pr1; Site_1: EcoRI;
             Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
             bacterial alkaline phosphatase (BAP) and then decapped
             with tabacco acid pyrophosphatase (TAP). The decapped
             intact mRNA was ligated with DNA-RNA linker including EcoR
             I site by treatment of T4 RNA ligase and the first strand
             cDNA was synthesized from oligo dT-selected mRNA by
             priming with dT-tailed vector. The dT-tailed vector was
             adjusted to have about 60nt. The cDNA vector was
             circularized with E. coli DNA ligase after digestion of
             EcoRI which site is also included in vector. An RNA strand
             converted to a DNA strand by Okayama-Berg method. The
             obtained cDNA vectors were used for transformation of
             competent cells E. coli Top10F by electroporation method.
             The cDNA libraries constructed by this method are
             full-length enriched cDNA library."
BASE COUNT   166 a      143 c      204 g      163 t
ORIGIN
Query Match      3.9%; Score 156; DB 14; Length 676;
Best Local Similarity 58.8%; Pred. No. 9.8e-35;

```

Matches	328: Conservative	0: Mismatches	220: Indels	10: Gaps
QY	879	TATTGGTGTGATGTTGGGATCGAGCATGTCTCCGCCCGGATTTTGTATCTCAACGGATC	938	
Db	125	TGTGGGTGGGAGCTGTGGAGGAGGAGTGTCCGTGTAGCTGTGGAGCAAGAGTGGGT	184	
QY	939	TCTGGTATCCATGCGCCACGAAATAATCATCGACTACGGCGCGCAGCGGAGCGGGTGA	998	
Db	185	CCTGTGGCTTTTGTGACAGACCACCATTTAAGAAGTGGGAGCCCACTTACACCATTA	244	
QY	999	ACAGTCCAGCAGGAGATCTGGCAGCGGGTGTTCATGTATTCGTATATGCGCTACTCT	1058	
Db	245	GCAGTCTCCGAGGACATCTGGCGTGGCGTGTGTGTGCACAAAGAAAGT-----TGT	298	
QY	1059	GGCAGACGTTTGTGCACAAAGTGTGGCAGGCATCGGTTTTATGCCACTGTTCCTGTGT	1118	
Db	299	ACAAAGGATGATTTTAAACCAATTCGAGGACATTGGTGTGATGTGCACGTCTCTGTGT	358	
QY	1119	GGTACTGATTAATAAAGCGATGATCATTTGGCTGTACAGCCCGGAGAGGATGATCAAGCAAA	1178	
Db	359	TGTTTTGGATTAAGCAGTTTACACCATTTACCAATCAACAGAAAGGAGATTCATCGA	418	
QY	1179	TATCATTTGTGTGATGATGATCACCGCGCCACCGCAAGCAGCAGATCAATGCCACTCA	1238	
Db	419	CGTCATTCATGTGGCTGGAGACATCGAGCACTAGTCAAGTTAACAGATCAATAGAGACCA	478	
QY	1239	CCATCCGGTGTCTGAACTACGTGCTGTGTAA--ATTTCGCTGAATGAAACCGCAAA	1297	
Db	479	GCACAGATGCTCCACAGTACGTGGGGGGGTGATGTCTGTGAATAATGACGCGCCGAAAC	538	
QY	1298	TTCCTCGGTGAAAGAAATATGCCAGAT---CTACGACAGTGTCCGGACATTTTGTG	1354	
Db	539	TTCCTGTGTGAAAGAACTTGAGAGATTTGCTGGATTAAGCGGGACATTTCTTTTG	598	
QY	1355	ATCTGAGCCGATTTTGTGACCTGGCGGAGCTACCGATGATTTACGCGGTTAGATGACACTG	1414	
Db	599	ATCTCCAGACTCTTATATGTGTGGAAGGCAACAGGTGTCAACACAGCGTCTCTGCTCC	658	
QY	1415	TTACCTGTAAATGACGT 1432		
Db	659	TGCTGTGTAAAGTGACAT 676		

RESULT 8
A2302589/c

LOCUS
DEFINITION
GSSBnu1525 *Brucella abortus* random genomic library *Brucella*
mellensis b1ovar Abortus genomic clone U01525, DNA sequence.

ACCESSION
A2302589

VERSION
A2302589.1

KEYWORDS
GI:10128800

SOURCE
GSS.

ORGANISM
Brucella mellensis b1ovar Abortus.
Brucella mellensis b1ovar Abortus.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; *Brucella*.
1 (bases 1 to 353)

REFERENCE
AUTHORS
Sanchez,D.O., Zaidomoni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
Facchi,P., Diaz,G., Lanzavecchia,S., Agnoro,F., Frasch,A.C.C.,
Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugaldé,R.A.
Gene discovery through genomic sequencing of *Brucella abortus*
Infect. Immun. 69 (2), 865-868 (2001)

TITLE
JOURNAL
MEDLINE
COMMENT
21101034

Contact: Siv Andersson
Small Genomes Sequencing Group
Department of Molecular Evolution, Uppsala University
Notbyvägen 18C, S-752 36, Uppsala, Sweden
Tel: 46-18-471-4379
Fax: 46-18-471-6404
Email: Siv.Andersson@ebc.uu.se

Sequences were basecalled with phred and vector was masked with
crossmatch (see <http://genome.washington.edu>). Sequences were then
trimmed from both ends to high quality bases and masked
vector.

Class: shotgun.
Location/Qualifiers
1. .535
/organism="Brucella melitensis biovar Abortus"
/strain="2308"
/db_xref="taxon:235"
/clone="U01525"
/clone_id="Brucella abortus random genomic library"
/note="Vector: modified M13"

BASE COUNT 123 a 174 c 144 g 94 t
ORIGIN

Query Match 3.9%; Score 153.8; DB 17; Length 535;
Best Local Similarity 57.1%; Pct. Ident. 34.4%;
Matches 300; Conservative 0; Mismatches 222; Indels 3; Gaps 1;

1716 CATGCAATCTACCACTTCTCCCTCGTTTGTACCGGGTGTCTGGGACCGTATTACAGTCC 1775
|||||
535 CATGACCAACCATGAAACAGCCGTTTTCGTGATGATGATGTCCTGGGGCCCTATTTCCTCGCC 476
1776 GATGCTCCAGGCTCTGTGTGTTAGTTGAAGCGCGGCAAAAGTCTCGCGGAGCAGTATTGA 1835
475 CATGTGCGCGGGCTGTGTGTTGAAGAAAGCGCGCAACATGCTGCGGGGCGCGTATTGA 416
1836 CCAGCTACTGTGATTTCCATCCGGCGCTGTTGAABAACCTCGCGAAATGGCAACACGTGTGA 1895
415 TCATCTGATTCACATGACATCCCTCGCGCGGGAAGCCGAAAGGACGCGCGGATCAAG 356
1896 TCAGCCCGCCCGCCGCTGTGCTGTGATCGATCGATTCCTCGAAAAAACGGCGCAACATCAGA 1955
355 CAAGGGGGCTGGCAGACAGCCCTCGTCGCCAA---GTTGAAGCTAAGAGTGTCTGAAAA 299
1956 TGCCTGCGCCCTGGCGGAAAGGGGCTACAGCTGTGCGCGGAATTTCTCGAAATCGCGCC 2015
298 AACGGCATGATCGTTGGGACATTCATGTCGTGCGGGAATTCCTCGGTAATCGTGACCC 239
2016 CTTCGCAATCTCTATGCGCAGAGCGGTAATTTGTGTGCTGGCTGTATGAGACGAGATCTGA 2075
238 TTTCCCGCCGCTCGATGCGCGCGGCTTAATTCGCGGGCGAGCACTTGATACAGGATGGA 179
2076 TAATTACTGCGCTGTATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2135
178 CAGTCTGGCGGACACTTATCTGCGCGGCTTTCGCGGCTGTGCTGATGATGATGATGATGAT 119
2136 TCTGACAGCTCAACAGCGGAGGAGTGAAGTAAATATCTTATTAAGCGGGGTGC 2195
118 TATCGAAGGCGAGCGGGCCAAAGGATATGTGTCTCACTATTTGCTCTCGCGCGCC 59
2196 CGGCGACGATCCACTGCTAGCAGCAATTTCTGGCAATACCTTCGG 2240
58 GCGCGGCTCAATCTCGTGGCGGACAGTCTCGCCGATGGAGCCG 14

RESULT 9
LOCUS AZ650403/c 655 bp DNA linear GSS 14-DEC-2000
DEFINITION J100520J01R Mouse 10kb plasmid U00C1M library Mus musculus genomic clone U00G100520J01 R, DNA sequence.
ACCESSION AZ650403
VERSION AZ650403.1 GI:11784856
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
1 (bases 1 to 655)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0520 row: L column: 01
Seq primer: CACACGAGAAACGCTATACC
Class: plasmid ends
High quality sequence stop: 655.

FEATURES

source

1. 655
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0520L01"
/clone.lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab.host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g147321419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 129 a 232 c 178 g 116 t

Query Match 3.8%; Score 150; DB 17; Length 655;

Best Local Similarity 54.0%; Pred. No. 6e-33;

Matches 353; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

OY 1024 GCGGCTGTTCATGATGTCGTAATGCGCTCCTGCGACAGCTTTGTGCACAAAGTGTG 1083
|||||
DB 655 GCGGCTGTTCATGATGTCGTAATGCGCTCCTGCGACAGCTTTGTGCACAAAGTGTG 596
OY 1084 GCGGATCGGCTTTGATGCACTGCTCTGTGTGTGCTACTGATATAAAGCGTATCCA 1143
|||||
DB 595 GCGGATCGGCTTTGATGCACTGCTCTGTGTGTGCTACTGATATAAAGCGTATCCA 536
OY 1144 TTGCGTGTGAGCGCGAAGAGATGCAAAACAATATCATTTGTGTGATGATGATCAACCG 1203
|||||
DB 535 GTTTCGCTCTCCGTCACTGCAAAAGACAGTGTGACACAGATCTGTGGCTTGACACCGC 476
OY 1204 GCCACGCAACAGCAGAGCAATCAATGCCACTACCATCGGTGCTGTAAGTACGTCGCT 1263
|||||
DB 475 GCATATGTCGAGCGCATGCTGACGCGATCGCGACCGGGTTTGGATTTTGGCGGT 416
OY 1264 GGTAAATTTTGGCTTAATGGAACACGAAATTTCTGTGGCTGAAAGAAATATGCCA 1323
|||||
DB 415 AACAGGCTGTCCGCGAAGATGAGATGCCAAGCTGATGCTGGAAGAAATTTGCGC 356
OY 1324 GAGATCTAGACAGTCCCGACAATTTTGGATCTGGCGGATTTTGGATCTGGCGGCT 1383
|||||
DB 355 GCAAGCTGGGCGCGCATGTCTTTCGCTTCGATCTGGCGGATTTTCCCTCACCTGGAAGCG 296
OY 1384 ACCGGTGAATTTAGCGGCTTCAGTATGCTACCTGTAAATGAGACGTTGGTGACAT 1443

DB 295 ACCGTTCCGGCCGAGCGTTCAATTCACGACGACGCGAAGTGAAATTTCTGGCGCAG 236
OY 1444 GAAATTCGCTGGGATCCAGATTTATTCGACCATTTGGCTTCACAGTATGAGTGA 1503
|||||
DB 235 GAAATTCGCGGTTGGCAGCGGAGGATTCATC -GAACTGGCGCGGCTTGTGATCTGAAGG 177
OY 1504 GATTTATTCGATTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1563
DB 176 ACAGGCGCGGCTGCGGACACGCGTATGCGCGGGAAGCATGCTCCGCTG- 121
OY 1564 ACAGCACAAGCCGCGGACAGATGGATTAATTCGCGCACACCTGTGCTGATGAGTTA 1623
DB 120 -CGCTGGAAGCGCGGACAGAACTTGCTGATACGAGGTTGCCAGTGGCGGCGGATG 62
OY 1624 ATTGATGCTTCACGCTGTGTGATCGATCGATGCGGTTGGCGGTAGAGGTGAGCGCTG 1677
DB 61 ATGATGCTTCATTCGCTGCTGCTGCGCGCGGCTTGGCGGATGCTTGGCGGATG 8

RESULT 10
BI557944 712 bp mRNA linear EST 05-SEP-2001
LOCUS 603237561F1.NC1.CGAP.Mam3 Mus musculus cDNA clone IMAGE:5290478 5',
DEFINITION mRNA sequence.

ACCESSION BI557944
VERSION BI557944.1 GI:15445258
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 712)
AUTHORS NIH-MGC, http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: c9abbs-r@mail.nih.gov
Tissue Procurement: Lotmar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LLNL1734 row: 3 column: 15
High quality sequence stop: 710.

FEATURES

source

1. 712
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5290478"
/clone.lib="NC1.CGAP.Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab.host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site.1: Salt; Site.2: Notti; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigators providing samples: Lotmar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT 174 a 152 c 211 g 175 t

Query Match 3.8%; Score 150; DB 13; Length 712;

Best Local Similarity 56.5%; Pred. No. 6.4e-33;

Matches 323; Conservative 0; Mismatches 240; Indels 9; Gaps 2;

OY 879 TATTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 938
|||||
DB 100 TGTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 159

QY 939 TGTGCTATCCATGCGCACAGAAAAATCAGACTACGGCGGCGAGCGGAAGCCGCTGGA 998
160 CATTGGGCTTTGGCAGACAGACCAATTAAAGTGGAGCCCTCAATTCAACCCACGCA 219
QY 999 ACAGTCCAGCAGAGATCTGGCAGCGGCTGTCTATCTATTCGTATATCGCTCACTCT 1058
220 GCAGTCCCTCAGAGGATATCTGGCGGCGCATGCTGCC-----TTGTCAAAAGGAAGTTGT 273
QY 1059 GAGAGAGCTTGTGTGACAAAGTGTGGCAGGCATCGGTTTGTATGACCCACTGTTCTGTGT 1118
274 TCAAGGATTTGATGACCATGATCGAGCATCTGTTTGTATGACCAAGTGTCTGTGT 333
QY 1119 GGTACTGGATTAAGAGCGTGTATCCATTCCTGTACGCCCGGAAGAGATCAAGCAAAA 1178
334 TGTCTTGAGTAAGAGATTCATCTCTTACCAGTAACCATGAAGGGAGCTCTCTCGAAA 393
QY 1179 TATCATTTGTGTGATGATACCGCGCCACCGAACAGCAGAGCAATCATGCTCACTCA 1238
394 CGTTATCATGTGTGCTGACACCGGCTGTACACGAGTCCACAGATTAATGAGACCAA 453
QY 1239 CCATCCGGTCTGACACTACGTGCTGTAATAATTTGCGCTGAATGGAACACCGAAT 1298
454 GCACAGAGTCTTCAAGATGTGGTGTGTATGTCTGTGAGATGCAAGCTCCAAAGCT 513
QY 1299 TCTCTGGCTGAAGAAATATGCGCAGAGAT--CTACGAACGTGCGCGACATTTTTCGA 1355
514 CCGTGGCTGAAGAGAACTTGAAGAGATTTGCTGGGATTAAGCGGAGACATTTCTTGA 573
QY 1356 TCTGGCGGATTTCTGACCTGGGCGGCTACCGGTGATTTAGCGGCTTCAATGACACTGT 1415
574 TCTCCAGAGACTCTTATCATGAGGAAGCAACAGAGTGTCTACAGCAGGTCTCTGTCTCTT 633
QY 1416 TACCTGTAAATGACGCTGCTGCGACATGAAA 1447
634 GGTGTCAAAATGACATACACAGACAGCAAAA 665
RESULT 11
B1149014 714 bp mRNA 1linear EST 05-JUL-2001
LOCUS B1149014
DEFINITION 602911220F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5052317 5',
mRNA sequence.
ACCESSION B1149014
VERSION B1149014.1 GI:14609015
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 714)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1142 row: 9 column: 06
High quality sequence start: 7
High quality sequence stop: 702.
Location/Qualifiers
1..714
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5052317"
/clone_11b="NCI_CGAP_L19"

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 175 a 153 c 219 g 167 t
ORIGIN
Query Match 3.8%; Score 149.2; DB 13; Length 714;
Best Local Similarity 56.2%; Pred. No. 1,1e-32;
Matches 324; Conservative 0; Mismatches 243; Indels 9; Gaps 2;
QY 879 TATTGGTGTGATGTGGGATCAGCAGCTGTCCGCGCGGATTTTGTATCTCAACGATC 938
121 TGTGGGATGATGTGGGATGTGGGATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 180
QY 939 TCTGCTATCCCATGCCACAAAAAATCAGACTACGCGCGGAGCGGAGCCGCTGGA 998
181 CTTTGTGGCTTTGGCAGAGCAGCAATTAAGAGTGGAGCTCAATTCACACACGCA 240
QY 999 ACAGTCCAGCAGAGATCTGGCAGGCGGTCTGTCTATGATTCGTATGCTGCTCACTCT 1058
241 GCAGTCTCAGAGGATATCTGGGCGCATGCTGCC-----TTGTCAAAAGGAAGTTGT 294
QY 1059 GGCAGAGCTTTGTGCACAAAGTGTGGCAGCATCGGTTTGTATGACCACTGTTCTGTGT 1118
295 TCAAGGATTTGATGACCATCGAATCCGAGCATTTGGTTTGTATGACCAAGTCTCTGTGT 354
QY 1119 GGTACTGGATTAAGAGCGTGTATCCATTCCTGTACGCCCGGAAGAGATCAAGCAAAA 1178
355 TGTCTTGAGTAAGAGATTCATCTCTTACCGATTAACCAAGAGGAGACCTCTCTGAAA 414
QY 1179 TATCATTTGTGTGATGATACCGCGCCACCGAACAGCAGAGCAATCATGCTCACTCA 1238
415 CGTTATCATGTGTGCTGACACCGGCTGTACACGAGTCCACAGATTAATGAGACCAA 474
QY 1239 CCATCCGGTCTGACACTACGTGCTGTAATAATTTGCGCTGAATGGAACACCGAAT 1298
475 GCACAGAGTCTTCAAGATGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 534
QY 1299 TCTCTGGCTGAAGAAATATGCGCAGAGAT--CTACGAACGTGCGCGACATTTTTCGA 1355
535 CCGTGGCTGAAGAGAACTTGAAGAGATTTGCTGGGATTAAGCGGAGACATTTCTTGA 594
QY 1356 TCTGGCGGATTTCTGACCTGGGCGGCTACCGGTGATTTAGCGGCTTCAATGACACTGT 1415
595 TCTCCAGAGCTCTTATCATGAGGAAGCAACAGGTGTACAGCAGCGGTCTGTCTGTCTT 654
QY 1416 TACCTGTAAATGACGCTGCTGCGACATGAAAATCG 1451
655 GGTGTCAAAATGACATACACAGACAGCAAAAAGTTG 690
RESULT 12
BF440790 503 bp mRNA 1linear EST 01-DEC-2000
LOCUS BF440790
DEFINITION 256813 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF440790
VERSION BF440790.1 GI:11500703
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
1 (bases 1 to 503)
TITLE Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grose, W.W., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
JOURNAL Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center

Db 357 TGTCTGGATTAAGAGTTCATCCTTTACAGTAACCAATGAGGGAGCTCTCTCGAAA 416
 QY 1179 TATCATTTGTGATGATGATCAACCGCCAGCAACAGCAATCAATGACATCA 1238
 Db 417 CGTTATCATGATGGCGGACACCGGGCTGTGACAGAGTTCACAGATTAATGACCAA 476
 QY 1239 CCATTCGGTCTGACATGATGCTGGTGTAAATTTCCCTGAATGAAACACGAAAT 1298
 Db 477 GCACAGAGTCTGATGATGCTGGTGTGTGATGCTGTGAGATGACAGCTCCAAAGCT 536
 QY 1299 TCTCTGGTGAAGAAATATGACAGAT---CTAGCAGGTGCCGACATTTTTCGA 1355
 Db 537 CCTGTGGTGAAGAGAGATGATGAGATTTGCTGGGATGAGCGGAGCAATTCCTTGA 596
 QY 1356 TCTGCGCATTTTCTGACCTGGCGGCTACCGGATGATTTAGCGGCTTCAGTATGCACTGT 1415
 Db 597 TCTCCAGATCTTCTATCATGATGAAGGCAACAGGTGTACACACAGGCTCTCTGCTCTT 656
 QY 1416 TACCTGTAATGACGT 1432
 Db 657 GGTGTGCAAAATGACAT 673

RESULT 14
 BM646051 670 bp mRNA linear EST 26-FEB-2002
 DEFINITION 1700087317822 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
 19600449631795 5', mRNA sequence.

ACCESSION BM646051
 VERSION BM646051.1 GI:18945562
 KEYWORDS EST.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anophelinae.

REFERENCE 1 (bases 1 to 670)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 TITLE 'R', Collins,F.H., Venter,J.C. and Hoffman,S.L.
 JOURNAL Celera Anopheles gambiae EST project
 COMMENT Unpublished (2002)
 CONTACT: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel.: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU010049M1 row: P column: 01
 Seq primer: M13 Reverse.

FEATURES
 source location/Qualifiers
 1..670

/organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449631795"
 /clone_1id="A.Gam.ad.cDNA1"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /note="Vector: pSport1; site_1: salI; site_2: NotI; whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.
 cDNA inserts >500 bp cloned directionally into pSport 1.
 Not 1 site is 3'. Clones available through the Malaria
 Research and Reference Reagent Resource Center
 (www.malaria.mf4.org)."
 BASE COUNT 146 a 184 c 212 g 128 t

Query Match 3.6%; Score 144.4; DB 13; Length 670;
 Best Local Similarity 56.0%; Pred. No. 2.9e-31;
 Matches 337; Conservative 0; Mismatches 256; Indels 9; Gaps 3;

QY 1086 AGGCATCGGTTTTCATCCACCTGTCTCTGGTGCTACTGGATTAAGAAACGGTATCCATT 1145

Db 32 AGGCATCGGTTTTCATCCACCTGTCTCTGGTGCTACTGGATTAAGAAACGGTATCCATT 91
 QY 1146 GCCTGTACCGCCGGAAGAGATGCAAGCA---AATATCATTTGGATGATGATCAACCG 1202
 Db 92 AAGGCTGCCCCGTTCTGGAAGACAAACGACGCAAGTATCTCTGGATGAGCAATCG 151
 QY 1203 CGCCACGCAACACAGACGCAATCAATGACCTACCATCCGCTGTAATCACTACGCG 1262
 Db 152 GCGGAGGAGAGGACAGGCTTCATCAACGCAACGACACCAACCAATGCTCAACTACGTCG 211
 QY 1263 TGTGTAATTTTCGCTGTAATGAAACACGCAAAATTTCTGCTGTAAGAAATATGCC 1332
 Db 212 TGGCAGCATTTTCGCTGCAATGAAAGTCCCAACCTCTGTGGCTCAAGCGAATGCA 271
 QY 1323 AGA---GATCTACGACAGTGGCGCAATTTTTCGATCTGGCGATTTTTCGACCTGGCG 1379
 Db 272 CGACGCGGTGTGACGAGAGTGGGGCATTTTACGACCTTCCGATTAACCTGACATTCGG 331
 QY 1380 GGCCTACCGGTGATTTAGCGGCTTCAGTATGACATGTTAATGAGAGTGGCTGGC 1439
 Db 332 TGCCACTGTGCGCATTTGCGGATTTGCTGCGCGCTGCAAAATGGAATCAATGACATGC 391
 QY 1440 ACATGAAATGCGTGGCATCCAGATTTATTTCCGACCATTTG---CCTTGACAGTTAGC 1496
 Db 392 AGAGATTCGATGCTGCTGCGAAGATTTCTGCAACGATCGTTTGGTGTGATGATCAAC 451
 QY 1497 GGATGAAGATTTTATTTGCTGATGCTCATATTTGTTTCCCGAACCCTTGCGAATA 1556
 Db 452 CGTGACACAGCTGGCGGCTGATCGGTACGACGTGGCAACCCGGGTGACCGATGCAACG 511
 QY 1557 TGTGTTAACACGACACAGCGCGGCAAGATGGATTTACTCCCGGACACCTGTGCTGT 1616
 Db 512 AGGACTTTCGGCGGACAGCGGCCCAAGAGCTGGGTTTAAACCTGCGACGCGATGAGCAAG 571
 QY 1617 AGGTTAATGATGCTACACGCTGTGGCATCGGTAGCAGTGGGCTTAAGAGTGAAGGCT 1676
 Db 572 TTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 631
 QY 1677 GA 1678
 Db 632 GA 633

RESULT 15
 BQ932942 983 bp mRNA linear EST 21-AUG-2002
 LOCUS BQ932942
 DEFINITION AGENCOURT_8782772 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374458
 5', mRNA sequence.

ACCESSION BQ932942
 VERSION BQ932942.1 GI:22348325
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 983)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@rs-remail.nih.gov
 Tissue Procurement: DCTD/DTF/gazdar
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LICM2553 row: h column: 11
 High quality sequence stop: 530.

FEATURES
 source location/Qualifiers
 1..983

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:6374458"  
/clone.lib="NIH_MGC_18"  
/tissue_type="large cell carcinoma"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pOT7; Site_1: XhoI; Site_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH_MGC Library."
```

```
BASE COUNT      235 a      234 c      284 g      230 t  
ORIGIN
```

```
Query Match      3.6%; Score 142.4; DB 14; Length 983;  
Best Local Similarity 58.8%; Pred. No. 1.5e-30;  
Matches 267; Conservative 0; Mismatches 181; Indels 6; Gaps 1;
```

```
OY 879 TATTGGTGTGATGGGATCAGGAGTGTCCGCGGAGATTTTGATCTCAACGATC 938  
Db 163 TGTGGGTGTGACGTTGGACAGGAGTGTCCGTGAGCTGTGTGACGAGTGGGT 222  
OY 939 TCTGTATCCATGCCACAGAAAAATCAGACTACGCGGCGAGCGAAGCCGCTGA 998  
Db 223 CCGTGTGGCTTTTGACAGACCAACCAATTAAGATTGGAGCCCACTTCAACCATGA 282  
OY 999 ACAGTCCAGCCAGAGATCTGGAGCGGTCTGTTCATGTATTGTAATGCCCTCACTCT 1058  
Db 283 GCAGTCTCCGAGACATCTGGGCTGTGTGTGTCAACAAGAAAGT-----TGT 336  
OY 1059 GGCAGACGTTTGTGCACAAAGTGTGGCAGCATCGTTTGATGCCACTGTCTGTGT 1118  
Db 337 ACAAGGATTTGATTAAACCAAAATCGAGGACTTGGTTGATGCCACGTTCTCTGTGT 396  
OY 1119 GGTACTGATTAATAACCGTGTATTCCTGTACGCCGGAAGAGATGCAAAAGAAA 1178  
Db 397 TGTTTTGATTAACGATTTACCCATTACAGTCAACCAAGAGGATTCCTCATGAAA 456  
OY 1179 TATCATTTGTGTGATGGATCACCAGCCGACGAAACAAGAGCGAATCAATGCCACTCA 1238  
Db 457 CGTCATCATGTGTGGTGCAGCCATCGAGTCAAGTTAACAGGATCAATGAGACCAA 516  
OY 1239 CCATCCGGTCTGATACGTGCGTGTGTAAATTTGCGCTGAATGGAACCGAATAT 1298  
Db 517 GCACAGTGTCTCCAGTACGTGGGGGGGTGATGTCTGTGAAATGCCGCCGAAACT 576  
OY 1299 TCTGTGGCTGAAGAAAAATATGCCAGATCTAC 1332  
Db 577 TCTGTGGCTGAAGAACTTGAGAGGATTTGC 610
```

```
Search completed: March 21, 2003, 23:49:25  
Job time : 6112.69 secs
```


GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 : Search time 20.4884 Seconds
(without alignments) 1625.927 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHNSVPSMNTPLNGKVAAL.....TRSKNVYRDVLVPGSVDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq_101002:*

1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	22	AAE09779
2	366	28.9	247	16	AAE07866
3	366	28.9	248	22	AAE07914
4	358	28.3	240	16	AAE07914
5	342	27.0	251	21	AAE55986
6	319.5	25.3	231	23	ABP39552
7	311	24.6	337	22	ABP10190
8	311	24.6	337	22	AAU23036
9	311	24.6	337	22	AAU18432
10	311	24.6	337	22	AAU42410

11	309	24.4	310	21	AAE27650	Human protein PRO2
12	309	24.4	310	22	AAE80238	Human PRO238 prote
13	309	24.4	310	23	ABE95441	Human angiotensin
14	309	24.4	310	23	ABE84835	Human PRO238 prote
15	309	24.4	325	21	AAE92513	Human OXRE-10. Ho
16	309	24.4	325	21	AAE79426	Human secreted pro
17	309	24.4	325	22	AAE38738	Human polypeptide
18	309	24.4	327	22	AAU18303	Human polypeptide
19	309	24.4	327	22	AAU40524	Human endocrine po
20	309	24.4	350	19	AAW74861	Human polypeptide
21	303	24.0	310	20	AAV13370	Human secreted pro
22	292	23.1	248	23	AAE47755	Human acid sequenc
23	291.5	23.0	254	22	AAE47522	Listeria monocytog
24	285	22.5	247	23	ABE48892	(R)-2-octanol dehy
25	285	22.5	248	21	AAE54422	Listeria monocytog
26	281.5	22.3	269	23	ABE92448	Amino acid sequenc
27	281.5	22.3	280	23	ABE92449	Herbicidally activ
28	278	22.0	206	23	AAE93345	Short-chain dehydr
29	275.5	21.8	253	23	ABE54214	Lactococcus lactis
30	275.5	21.8	272	21	AAE54420	Secoisolaricresin
31	275.5	21.8	273	21	AAE54412	Secoisolaricresin
32	273	21.6	202	23	AAU77210	Consensus protein
33	272.5	21.5	277	21	AAE54415	Secoisolaricresin
34	271	21.4	260	21	AAE51581	Arabidopsis thalia
35	271	21.4	260	23	ABE92475	Arabidopsis thalia
36	271	21.4	261	21	AAE24387	Arabidopsis thalia
37	271	21.4	274	21	AAE51580	Arabidopsis thalia
38	271	21.4	276	21	AAE24386	Arabidopsis thalia
39	271	21.4	286	21	AAE51575	Arabidopsis thalia
40	270.5	21.4	251	22	AAU34193	Staphylococcus aur
41	270.5	21.4	272	22	AAU37095	Staphylococcus aur
42	267.5	21.1	277	21	AAE54413	Secoisolaricresin
43	265	20.9	250	23	ABE54881	Lactococcus lactis
44	262.5	20.8	241	22	ABE96397	Putative P. abyssi
45	261	20.6	249	23	ABE93980	Staphylococcus epi

ALIGNMENTS

RESULT 1	AAE09779	standard; Protein; 250 AA.
ID	AAE09779	
XX	AAE09779;	
AC	AAE09779;	
XX	29-NOV-2001 (first entry)	
DE	Escherichia coli strain C rfl operon encoding ribitol dehydrogenase.	
XX	Positive selection system: metabolise: arabitol; ribitol; mannitol;	
KW	transgenic cell; marker gene; ribitol dehydrogenase; ribitol kinase;	
KW	ribitol transporter; rfl operon.	
XX		
OS	Escherichia coli C.	
XX		
PN	W0200166779-A2.	
PD	13-SEP-2001.	
XX		
PF	08-MAR-2001; 2001WO-US07474.	
XX		
PR	08-MAR-2000; 2000US-0188291.	
XX	15-AUG-2000; 2000US-0255595.	
PI	(UYGE-) UNIV GEORGIA RES FOUND INC.	
XX	Parrott W, Lafayette P, Kane P;	
XX	WPI: 2001-565596/63.	
DR	N-PSDB; AAD16811.	
XX		
PT	Positively selecting transformed cells comprising selectable marker	

PT gene and desired gene, from a cell population by using marker compounds
 PT e.g., arabinol, ribitol which confer selective advantage on transformed
 PT cells
 XX
 PS Claim 16: Page 37; 37pp; English.
 CC The present invention relates to a positive selection system that
 CC involves conferring to transferred cells the ability to metabolise
 CC arabinol, ribitol and/or mannitol. The positive selection method is
 CC used in positively selecting transgenic cells from a population of cells
 CC using the positive selection method, the presence of the gene of
 CC interest in the genetically transformed cells may be determined without
 CC the disadvantages associated with traditional negative selection
 CC systems. Positive selection of the transformed cells is achieved without
 CC directly damaging the neighbouring non-transformed cells. The
 CC transformed cells may be identified by simple visual means without the
 CC use of a separate assay to determine the presence of a marker gene. This
 CC technique also avoids the release of antibiotics or other dangerous
 CC genes into the environment. The present sequence is Escherichia coli
 CC strain C ribitol dehydrogenase encoded by ribitol operon (rbl operon).
 CC The operon also encodes ribitol kinase and ribitol transporter.
 CC
 XX
 SQ Sequence 250 AA;
 Query Match 100.0%; Score 1265; DB 22; Length 250;
 Best Local Similarity 100.0%; Pred. No. 1.4e-127;
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MMNHSVSMPTPLNGKVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 60
 Db 1 MMNHSVSMPTPLNGKVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 60
 QY 61 ENAVAIQLDLFNNQVDNMLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVNLN 120
 Db 61 ENAVAIQLDLFNNQVDNMLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVNLN 120
 QY 121 INNAFRCVAVLPKMIQORSGDITFTSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVS 180
 Db 121 INNAFRCVAVLPKMIQORSGDITFTSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVS 180
 QY 181 QYGVRAVAVLPGPVVTALLDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 240
 Db 181 QYGVRAVAVLPGPVVTALLDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 240
 QY 241 LVIIPGSVDL 250
 Db 241 LVIIPGSVDL 250
 RESULT 2
 AAR77866
 ID AAR77866 standard; Protein; 247 AA.
 AC AAR77866;
 XX
 DT 13-NOV-1995 (first entry)
 XX
 DE S. clavuligerus ORP9 product.
 XX
 KM Clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.
 XX
 OS Streptomyces clavuligerus.
 XX
 PN CA2108113-A.
 XX
 PD 09-APR-1995.
 XX
 PF 08-OCT-1993; 93CA-2108113.
 XX
 PR 08-OCT-1993; 93CA-2108113.
 XX
 PA (UYAL-) UNIV ALBERTA.
 XX

PI Aideo KA, Jensen SE, Paradkar AS;
 XX
 DR WPI: 1995-207301/28.
 DR N-PSDB: AAQ91580.
 XX
 PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
 PT biosynthesis of the antibiotic in Streptomyces hosts which do not
 PT naturally produce clavulinate
 XX
 PS Claim 31; Fig.18; 41pp; English.
 CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (AAQ91580),
 CC extending downstream from pcbc, included 10 ORFs encoding the
 CC enzymes required for clavulinate biosynthesis. ORF8 (oriented
 CC in the opposite direction to pcbc) encoded a protein (AAR77866)
 CC that showed weak similarity to ribitol 5 P04 dehydrogenase-type
 CC enzymes.
 CC
 XX
 SQ Sequence 247 AA;
 Query Match 28.9%; Score 366; DB 16; Length 247;
 Best Local Similarity 37.3%; Pred. No. 7e-31;
 Matches 91; Conservative 41; Mismatches 104; Indels 8; Gaps 3;
 QY 9 MNTPLNKRVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 65
 Db 1 MNTPLNKRVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 65
 QY 66 LQDLFNNQVDNMLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVNLINNAF 125
 Db 61 LELDVADROGVDAVASTVEALGSLDLVNNAGIMLGPEVDADPTTWTMTIDNLLGLM 120
 QY 126 RCVRAVLPKMIQORSGDITFTSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVSQYGR 185
 Db 121 YMTFRAALPHILRSK-GTVVQMSIAGRVNVRNNAVYQATKFGVNAFSETLROEVTGRGVR 179
 QY 186 VGAVLPBPVVTAL---LDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 241
 Db 180 VVVEPEPTTDELGHITTHATKEMPEORISQIRKIQADIAEAVRAYVAPPHATVHEI 239
 QY 242 VILP 245
 Db 240 FIRP 243
 RESULT 3
 AAEO7914
 ID AAEO7914 standard; Protein; 248 AA.
 AC AAEO7914;
 XX
 DT 01-NOV-2001 (first entry)
 XX
 DE S. clavuligerus clavulanic acid biosynthesis enzyme #6.
 XX
 KM Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
 KM broad spectrum beta-lactamase inhibitor; open reading frame; ORF;
 KW pcbc gene.
 XX
 OS Streptomyces clavuligerus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 247..248
 FT /note="Encoded by GTC on the inverse complementary
 strand of the corresponding DNAs (AAD14499, AAD14510)"
 XX
 PN US6232106-B1.
 XX
 PD 15-MAY-2001.
 XX
 PF 30-AUG-1999; 99US-0385028.
 XX
 PR 29-JAN-1997; 97US-0790462.
 XX

PR 08-OCT-1993; 93US-0134018.
PR 06-DEC-1995; 95US-0567801.
XX
PA (UYAL-) UNIV ALBERTA.
XX
PI Jensen SE, Aldoo KA, Paradkar AS;
XX
DR MPI: 2001-342772/36.
DR N-PSDB; AAD14499, AAD14510.
XX
PT Novel enzyme required for clavulanic acid biosynthesis which is useful
XX as broad spectrum beta-lactamase inhibitor .
XX
PS Claim 1; Fig 17; 75pp; English.
XX
CC The invention relates to DNA sequences encoding enzymes required
CC for clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum
CC beta-lactamase inhibitor and is an important antibiotic for the treatment
CC of infectious diseases. Also provided in the patent is a 15 kb genomic
CC DNA fragment downstream to pcbc gene from streptomyces clavuligerus.
CC The genomic DNA comprises 10 open reading frames (ORFs), eight of which
CC are involved in clavulanic acid biosynthesis. The present sequence is
CC S. clavuligerus clavulanic acid biosynthesis enzyme encoded by ORF9.
XX
SQ Sequence 248 AA;

Query Match 28.9%; Score 366; DB 22; Length 248;
Best Local Similarity 37.3%; Pred. No. 7e-31; Indels 8; Gaps 3;
Matches 91; Conservative 41; Mismatches 104;

OY 9 MNTPLNGKVAATITGAASGIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 65
DB 1 MSPALGKVALITIGASSGIGEGATRALAEGAAVAIAARVEKRLAGDELTAAGAKVHV 60
OY 66 LQIDLNFNNOQVDMNLADITELAGGLDIFHNACAYIGCPVAEGDPDWDVNLNINAAF 125
DB 61 LEIDVADRCGVDAVAVASTVEALGGLDILVNNAGIMLGPVEDADDTTMTMTIDITNLGLM 120
OY 126 RCRAVLPNMIARSGDIIFTSSIAGVVPYIMPTITASKFAVOAFVHTTRROVSQYGR 185
DB 121 YMRALPPLIRSK-GTVVOMSSIAGRVNNAVAVQATFGVNAFSETTLROEYTERGV 179
OY 186 VCAVLPGPVVATL---LDDMPKAMEALANGSLMOPFEVASVLEPMVTRSKNVTVRD 241
DB 180 VVVEIEPTDTTELRGHTHTATKEMYEQRISQIRKLOAQDIAEVRVAVAPRNHATVNEI 239
OY 242 VILP 245
DB 240 FIRP 243

RESULT 4
AAR61477
ID AAR61477 standard; Protein; 240 AA.
XX
AC AAR61477;
XX
DT 01-OCT-1995 (first entry)
XX
DE Clavulanic acid dehydrogenase sequence.
XX
KW Clavulanic acid; antibiotic; Augmentin.
XX
OS Streptomyces clavuligerus ATCC 27064.
XX
PN WO9503416-A.
XX
PD 02-FEB-1995.
XX
PF 15-JUL-1994; 94MO-EP02346.
XX
PR 24-JUL-1993; 93GB-0015393.
XX

PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Arnell J, Elson SW, Nicholson NH, Woroniecki SR;
XX
DR MPI: 1995-075242/10.
DR N-PSDB; AAO67531.
XX
PT New clavulanic acid dehydrogenase from Streptomyces, and related
XX DNA and vectors - used to produce beta-lactamase inhibiting
XX clavulanic acid from new 3-oxo:ethylidene analogues
XX
PS Claim 2; Fig 1; 36pp; English.

CC A new enzyme is disclosed which has clavulanic acid dehydrogenase
CC activity, the enzyme having an apparent mol. wt. of 28 kD (by SDS
CC PAGE) and incorporating the N-terminal sequence PSALGKVALITIGASSGIGL.
CC The enzyme is derived from the mycelium of a Streptomyces species,
CC especially S. clavuligerus (e.g. ATCC 27064), S. junoensis (e.g. ATCC
CC 29864) or S. katushahamans (e.g. T272). The present sequence
CC represents the enzyme from S. clavuligerus ATCC 27064.
CC The enzyme is used to catalyse the biosynthesis of clavulanic acid
CC from a precursor clavulanic acid aldehyde. The obtained clavulanic acid
CC is in turn a key ingredient in the antibiotic Augmentin.
XX
SQ Sequence 240 AA;

Query Match 28.3%; Score 358; DB 16; Length 240;
Best Local Similarity 37.1%; Pred. No. 4.9e-30; Indels 8; Gaps 3;
Matches 89; Conservative 41; Mismatches 102;

OY 9 MNTPLNGKVAATITGAASGIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 65
DB 1 MSPALGKVALITIGASSGIGEGATRALAEGAAVAIAARVEKRLAGDELTAAGAKVHV 60
OY 66 LQIDLNFNNOQVDMNLADITELAGGLDIFHNACAYIGCPVAEGDPDWDVNLNINAAF 125
DB 61 LEIDVADRCGVDAVAVASTVEALGGLDILVNNAGIMLGPVEDADDTTMTMTIDITNLGLM 120
OY 126 RCRAVLPNMIARSGDIIFTSSIAGVVPYIMPTITASKFAVOAFVHTTRROVSQYGR 185
DB 121 YMRALPPLIRSK-GTVVOMSSIAGRVNNAVAVQATFGVNAFSETTLROEYTERGV 179
OY 186 VCAVLPGPVVATL---LDDMPKAMEALANGSLMOPFEVASVLEPMVTRSKNVTVRD 241
DB 180 VVVEIEPTDTTELRGHTHTATKEMYEQRISQIRKLOAQDIAEVRVAVAPRNHATVNEI 239

RESULT 5
AAY55986
ID AAY55986 standard; Protein; 251 AA.
XX
AC AAY55986;
XX
DT 15-MAR-2000 (first entry)
XX
DE S.clavuligerus'cladh protein.
XX
KW Operon; cladh; clar; biosynthetic pathway; clavulanic acid; enzyme;
KW clavulanate-9-aldehyde reductase; transcriptional regulator;
KW gene expression.
XX
OS Streptomyces clavuligerus.
XX
PN ES2131001-A1.
XX
PD 01-JUL-1999.
XX
PF 16-JUN-1997; 97ES-0001305.
XX
PR 16-JUN-1997; 97ES-0001305.
XX
PA (ANTI) ANTIBIOTICOS SAU.
XX

DR	WPI: 2000-026016/03.
DR	N-PSDB: AAZ30700.
XX	
PT	Improving production of clavulanic acid by Streptomyces - by
PT	super-expression of clar gene.
XX	
PS	Example 1; Page 8-9; 23pp; Spanish.
XX	
CC	This sequence represents the clath protein which is a putative
CC	clavulanate-9-aldohyde reductase used in the biosynthetic pathway for
CC	clavulanic acid in Streptomyces clavuligerus. The invention relates to
CC	methods of improving production of clavulanic acid by Streptomyces by
CC	super-expression of clar gene. The clar gene is characterised in that
CC	it is localised in the gene group encoding genes for biosynthesis of
CC	clavulanic acid.
XX	
SO	Sequence 251 AA:
Query Match	27.0%; Score 342; DB 21; Length 251;
Best Local Similarity	35.7%; Pred. No. 2.7e-28;
Matches 87; Conservative 40; Mismatches 109; Indels 8; Gaps 3	
OY	9 MMTPLNGKVAATTTGAASGIGLOCATLLDAGAKVYLIDREGKLRKIYAE--GENATA 65
Db	5 MPSALQKVALTTGREGELGRRAATALAEAGAVALAARVVEKRLALGDELTAAGAKVAV 64
OY	66 LQLDLNNQOVNMMADIIETLAGDIFHANGAVATGCGVACGDPDVMRVNLNINAAF 125
Db	65 LELVDADNGQVAAVAASVTEALGGLDILVNNNGIMLGSVEADDDTTDMRIDTNLGLM 124
OY	126 KCVRAVLPHMTAQRSGDIIFTSSINGVVPVITWEPIYTASKFAVQAFVHTTRQVSQYGR 185
Db	125 YMTRALPHILRSK-GTVVQMSISAGRVYRNAAVYQATKFCGVNAFSETVROEVTTERGV 183
OY	186 VGAVLPGPVVTVTL---LDDMPKAMEEALANGSLMOPTEVAESVLEFVTRSKNTVTRDL 241
Db	184 VVIERPGTITDELRRGHITHTATKREYEQKISIRKLQADIDAEAVRYAVTAAHHATVHEI 243
OY	242 VILP 245
Db	244 FIRP 247
RESULT 6	
ABP39552	
ID	ABP39552 standard; Protein; 231 AA.
XX	ABP39552;
XX	
DT	24-JUL-2002 (first entry)
XX	
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4397;
XX	
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX	antibacterial; gene therapy.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	US6380370-B1.
XX	
PD	30-APR-2002.
XX	
PF	13-AUG-1998; 98US-0134001.
XX	
PR	14-AUG-1997; 97US-055779P.
XX	
PR	08-NOV-1997; 97US-064964P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Doucette-Stamm LA, Bush D;
XX	
XX	WPI: 2002-381255/41.
DR	N-PSDB: ABN92097.
DR	

[illegible]

PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225477P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226688P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 25-SEP-2000; 2000US-234984P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235835P.
PR 29-SEP-2000; 2000US-236377P.
PR 29-SEP-2000; 2000US-236378P.
PR 29-SEP-2000; 2000US-236379P.
PR 29-SEP-2000; 2000US-236380P.
PR 29-SEP-2000; 2000US-236381P.
PR 29-SEP-2000; 2000US-236382P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239937P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.

PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-25030P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251858P.
PR 08-DEC-2000; 2000US-251859P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.

(HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Barash SC, Ruben SM.

XX WPI; 2001-476161/51.

DR N-PSDB; ABA06412.

XX Isolated nucleic acid molecule encoding an inflammation-associated

PT polypeptide is used in preventing, treating or ameliorating a medical

PT condition

XX Claim 11; SEQ ID NO: 498; 859pp + Sequence Listing; English.

CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.

PR 07-JUL-2000: 2000US-0216647.
 PR 07-JUL-2000: 2000US-0216880.
 PR 11-JUL-2000: 2000US-0217487.
 PR 11-JUL-2000: 2000US-0217496.
 PR 14-JUL-2000: 2000US-0218290.
 PR 26-JUL-2000: 2000US-0220963.
 PR 26-JUL-2000: 2000US-0220964.
 PR 14-AUG-2000: 2000US-0224518.
 PR 14-AUG-2000: 2000US-0224519.
 PR 14-AUG-2000: 2000US-0225213.
 PR 14-AUG-2000: 2000US-0225214.
 PR 14-AUG-2000: 2000US-0225266.
 PR 14-AUG-2000: 2000US-0225267.
 PR 14-AUG-2000: 2000US-0225268.
 PR 14-AUG-2000: 2000US-0225270.
 PR 14-AUG-2000: 2000US-0225447.
 PR 14-AUG-2000: 2000US-0225757.
 PR 14-AUG-2000: 2000US-0225758.
 PR 14-AUG-2000: 2000US-0225759.
 PR 18-AUG-2000: 2000US-0226279.
 PR 22-AUG-2000: 2000US-0226681.
 PR 22-AUG-2000: 2000US-0226688.
 PR 23-AUG-2000: 2000US-0227182.
 PR 30-AUG-2000: 2000US-0228924.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229343.
 PR 01-SEP-2000: 2000US-0229344.
 PR 01-SEP-2000: 2000US-0229345.
 PR 05-SEP-2000: 2000US-0229509.
 PR 05-SEP-2000: 2000US-0229513.
 PR 06-SEP-2000: 2000US-0230437.
 PR 06-SEP-2000: 2000US-0230438.
 PR 08-SEP-2000: 2000US-0231242.
 PR 08-SEP-2000: 2000US-0231243.
 PR 08-SEP-2000: 2000US-0231244.
 PR 08-SEP-2000: 2000US-0231413.
 PR 08-SEP-2000: 2000US-0231414.
 PR 08-SEP-2000: 2000US-0232080.
 PR 08-SEP-2000: 2000US-0232081.
 PR 12-SEP-2000: 2000US-0231968.
 PR 14-SEP-2000: 2000US-0232397.
 PR 14-SEP-2000: 2000US-0232398.
 PR 14-SEP-2000: 2000US-0232399.
 PR 14-SEP-2000: 2000US-0232400.
 PR 14-SEP-2000: 2000US-0232401.
 PR 14-SEP-2000: 2000US-0233063.
 PR 14-SEP-2000: 2000US-0233064.
 PR 14-SEP-2000: 2000US-0233065.
 PR 21-SEP-2000: 2000US-0234223.
 PR 21-SEP-2000: 2000US-0234274.
 PR 25-SEP-2000: 2000US-0234997.
 PR 25-SEP-2000: 2000US-0234998.
 PR 26-SEP-2000: 2000US-0235484.
 PR 27-SEP-2000: 2000US-0235834.
 PR 27-SEP-2000: 2000US-0235836.
 PR 29-SEP-2000: 2000US-0236327.
 PR 29-SEP-2000: 2000US-0236367.
 PR 29-SEP-2000: 2000US-0236368.
 PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239335.
 PR 13-OCT-2000: 2000US-0239337.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241221.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.

PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.
 PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.
 PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249212.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249264.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251899.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM,
 WPI: 2001-451936/48.
 DR N-PSDB: AAS29661.
 PT Isolated polypeptide for treating, preventing and/or prognosing
 disorders of the endocrine system such as reproductive disorders,
 endocrine cancers and also for testing and detection e.g. diagnosis
 PS Claim 11: SEQ ID No 387; 604bp; English.
 CC Sequences AAU18282-AAU18507 represent endocrine polypeptides of the
 CC invention. Endocrine polypeptides and their associated polynucleotides
 CC are useful in the diagnosis, treatment and prevention of various types of
 CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
 CC chickens or sheep. A pathological condition can be determined by
 CC determining the presence or absence of a mutation in an endocrine
 CC polynucleotide. The treatable disorders include autoimmune diseases such

CC as rheumatoid arthritis, hyperproliferative disorders such as neoplasms
 CC of the breast or liver, cardiovascular disorders such as cardiac arrest,
 CC cerebrovascular disorders such as cerebral ischemia, nervous system
 CC disorders such as Alzheimer's disease, infections caused by bacteria,
 CC viruses and fungi, ocular disorders such as corneal infection, endocrine
 CC disorders such as premature labour and infertility, gastrointestinal
 CC disorders such as Crohn's disease, renal disorders such as
 CC glomerulonephritis and respiratory disorders such as asthma. The
 CC polypeptides can also be used to aid wound healing, to prevent skin aging
 CC due to sunburn, to maintain organs before transplantation, to regenerate
 CC tissues and in chemotaxis. The polypeptides can also be used as a food
 CC additive or preservative to increase or decrease storage capabilities.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match 24.6%; Score 311; DB 22; Length 337;
 Best Local Similarity 33.7%; Pred. No. 8,9e-25;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

OY 13 LNKVVAITGAASGIGLCAKTLTLDGAKVYLIDREGDKHKIVAEI-----GENA 63
 DB 62 LRNAVAVITGATSGLKECKKVFYAGAKLVLCGRNGALEELIRELTASHTKVOYTHKP 121
 OY 64 YALQDLFNNQVDNMLADIIELAGLIDIFHANAGAVTGGPVAECPDVWDVRYLNININA 123
 DB 122 YLVTPLDLSGALVAAAELQCGFYVDILVNNAGISYRTINDTYVDVKRMENNYEG 181
 OY 124 AFRCVAVLPHMIAORSGLDIFTSIAGVVPVIMEPIYTASKRAVOAFVHTTRQVSQYG 183
 DB 182 PVALTALPLSMIKRQGHVVAISSIOGKMSIFFRSAVYASKHATQAFECLEAEQOYE 241
 OY 184 VRVGAIVLPGPV-----VTALDMDMPK-AKMEALANGSLMOPLEVESVLFVY-TTSKNV 236
 DB 242 IEYTVISPGYIHTNLSVNAITTDGSKRYGVMDTTTAAQG--RSPVEVADVYLAAGKKKKDY 299
 OY 237 TVRDLY 242
 DB 300 ILADLL 305
 RESULT 10
 AAM42410
 ID AAM42410 standard; Protein; 337 AA.
 AC XX
 DT XX
 DE 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 143.
 KW Humen; nootropic; antiinflammatory; anti-HIV; antibacterial; virucide;
 KW immunosuppressive; antiinflammatory; antianaemic; antiarthritic; cancer;
 KW antiparkinsonian; antistroke; cerebroprotective; antiinflammatory;
 KW antirheumatic; hepatoprotective; antidiabetic; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 OS Homo sapiens.
 XX
 PN WO200155449-A1.
 PD 02-AUG-2001.
 PE 17-JAN-2001; 2001WO-US01346.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUL-2000; 2000US-0216880.
 PR 14-JUL-2000; 2000US-0218290.
 PR 14-AUG-2000; 2000US-0223447.

PR 01-SEP-2000; 2000US-0229343.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231243.
 PR 25-SEP-2000; 2000US-0234997.
 PR 29-SEP-2000; 2000US-0236367.
 PR 13-OCT-2000; 2000US-0239937.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249265.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-476225/51.
 DR N-PSDB; AAI62815.

PT Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders
 XX
 PS Claim 11; SEQ ID NO 143; 532pp + Sequence Listing; English.

XX The invention relates to novel genes (AA162752-AA162961) and proteins
 CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX

Sequence 337 AA;
 Query Match 24.6%; Score 311; DB 22; Length 337;
 Best Local Similarity 33.7%; Pred. No. 8,9e-25;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

OY 13 LNKVVAITGAASGIGLCAKTLTLDGAKVYLIDREGDKHKIVAEI-----GENA 63
 DB 62 LRNAVAVITGATSGLKECKKVFYAGAKLVLCGRNGALEELIRELTASHTKVOYTHKP 121
 OY 64 YALQDLFNNQVDNMLADIIELAGLIDIFHANAGAVTGGPVAECPDVWDVRYLNININA 123
 DB 122 YLVTPLDLSGALVAAAELQCGFYVDILVNNAGISYRTINDTYVDVKRMENNYEG 181
 OY 124 AFRCVAVLPHMIAORSGLDIFTSIAGVVPVIMEPIYTASKRAVOAFVHTTRQVSQYG 183

Db 182 PVALTALKLPSMIKROGHIVAISISIOGKMSIPFRSAVAASKHATQAFDCLRAEMEQYE 241
 QY 184 VRGAVLPGPV-----VTALIDDPK-AKMEALANGSLMOPIEVAESVLFMV-TRSKNV 236
 Db 242 IEVTVISPGYIHTNLISYNAITADGSRYGVDITTAQG--RSPVEVADVLAAGKKKKDV 299
 QY 237 TVRDV 242
 Db 300 ILADL 305

RESULT 11
 AAB27650
 ID AAB27650 standard; Protein; 310 AA.
 AC AAB27650;
 XX
 DT 26-JAN-2001 (first entry)
 XX
 DE Human protein PRO238.
 XX
 KW Cardiovascular; endothelial; angiogenic disorder; PRO179;
 KW PRO238; PRO364; PRO844; PRO1760; PRO205; PRO321; PRO333;
 KW PRO840; PRO877; PRO878; PRO879; PRO882; PRO885; PRO887;
 KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 XX /label= "Signal peptide"
 XX
 PN WO200053757-A2.
 PD 14-SEP-2000.
 XX
 PE 24-FEB-2000; 2000MO-US05004.
 XX
 PR 08-MAR-1999; 99MO-US05028.
 PR 12-MAR-1999; 99US-012357.
 PR 02-JUN-1999; 99MO-US12252.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99MO-US20111.
 PR 15-SEP-1999; 99MO-US21090.
 PR 30-NOV-1999; 99MO-US28313.
 PR 30-NOV-1999; 99MO-US28409.
 PR 02-DEC-1999; 99MO-US28565.
 PR 18-FEB-2000; 2000MO-US04342.
 PR 22-FEB-2000; 2000MO-US04414.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Gertsen ME;
 PI Goddard A, Gurney AL, Hillan KJ, Marsters SA, Peoni NF, Pitti RM;
 PI Wetanabe CK, Williams PM, Wood WT.
 XX
 DR WPI; 2000-611444/58.
 DR N-PSDB; AAA99902.
 XX
 PT Novel PRO polypeptides and agonists and antagonists of them, used to
 PT diagnose and treat cardiovascular, endothelial and angiogenic disorders
 PT
 PS -
 PS Claim 71; Fig 4; 181pp; English.
 XX
 CC The present invention relates to methods for stimulating or inhibiting
 CC angiogenesis and cardiovascularization. The methods involve the use of
 CC pharmaceutical compositions based on the following proteins, PRO179,
 CC PRO238, PRO364, PRO844, PRO846, PRO1760, PRO205, PRO321, PRO333,
 CC PRO840, PRO877, PRO878, PRO879, PRO882, PRO885 or PRO887. These
 CC proteins were identified by isolating cDNA clones encoding secreted
 CC proteins. The proteins of the invention may be used to diagnose and

CC treat cardiovascular, endothelial or angiogenic disorders. The present
 CC sequence is one of the proteins of the invention.
 XX
 SQ Sequence 310 AA;
 Query Match 24.4%; Score 309; DB 21; Length 310;
 Best Local Similarity 33.7%; Pred. No. 1.3e-24;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

QY 13 LNCRAVAAITGAASGIGLQCAKTLIDAGAKVYLIDREGDKLHKIVAEI-----GENA 63
 Db 35 LRNAVYVITGATSGLGECAKVFYAAGAKVLCGRNGALBELRELTAHATKVCQHRK 94
 QY 64 YALQDLFPNQVDNMLADILELAGLDIFHANAAYIGSPVAEGPDPVDRVLNLINA 123
 Db 95 YLVTFEDLTDSGATVAANAELIQCFCGYVDILVNNAGISYRGTIMDTVDVDRVETNYFG 154
 QY 124 AFRCVRAVLPHMIAQRSGDIIFTSSINGVVPVIMEPIYTAASKFAYQAFVHTTROVSQY 183
 Db 155 PVALTALKLPSMIKROGHIVAISISIOGKMSIPFRSAVAASKHATQAFDCLRAEMEQYE 214
 QY 184 VRGAVLPGPV-----VTALIDDPK-AKMEALANGSLMOPIEVAESVLFMV-TRSKNV 236
 Db 215 IEVTVISPGYIHTNLISYNAITADGSRYGVDITTAQG--RSPVEVADVLAAGKKKKDV 272
 QY 237 TVRDV 242
 Db 273 ILADL 278

RESULT 12
 AAB80238
 ID AAB80238 standard; Protein; 310 AA.
 XX
 AC AAB80238;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human PRO238 protein.
 XX
 KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
 KW antiParkinsonian nootropic; neuroprotective; vulnery; candid;
 KW antiangiogenic; vasotropic; antiasthmatic; antineuritic; cancer;
 KW antiarthritic; antidiabetic; antidiabetic; antidiabetic; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KW Ischaemia; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200104311-A1.
 PD 18-JAN-2001.
 XX
 PE 22-FEB-2000; 2000MO-US04414.
 XX
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99MO-US20594.
 PR 13-SEP-1999; 99MO-US20944.
 PR 15-SEP-1999; 99MO-US21090.
 PR 15-SEP-1999; 99MO-US21547.
 PR 05-OCT-1999; 99MO-US23089.
 PR 29-NOV-1999; 99MO-US28214.
 PR 30-NOV-1999; 99MO-US28313.
 PR 16-DEC-1999; 99MO-US30095.
 PR 20-DEC-1999; 99MO-US30911.
 PR 20-DEC-1999; 99MO-US30999.
 PR 05-JAN-2000; 99MO-US00219.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 6.82947 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHNSVPMNTPLNGKVAAL.....TRSKNVTRDLVILPGSYDL 250

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	366	28.9	248	US-09-385-028-11	Sequence 11, Appl
2	319.5	25.3	231	US-09-134-001C-4397	Sequence 4397, Ap
3	309	24.4	350	US-09-149-476-452	Sequence 452, App
4	261	20.6	249	US-09-134-001C-4825	Sequence 4825, Ap
5	255	20.2	246	US-09-238-481-2	Sequence 2, Appl
6	255	20.2	246	US-09-572-810A-2	Sequence 2, Appl
7	247	19.5	258	US-09-504-358-12	Sequence 12, Appl
8	247	19.5	258	US-09-954-314-12	Sequence 12, Appl
9	247	19.5	253	US-09-134-001C-4512	Sequence 12, Ap
10	246.5	19.5	263	US-09-287-097-2	Sequence 2, Appl
11	245.5	19.4	333	US-08-440-856A-4	Sequence 4, Appl
12	242.5	19.2	274	US-09-134-001C-4431	Sequence 4431, Ap
13	242	19.1	262	US-09-363-189B-6	Sequence 6, Appl
14	241	19.1	267	US-09-134-001C-5042	Sequence 5042, Ap
15	231.5	18.3	315	US-08-793-035-9	Sequence 9, Appl
16	231.5	18.3	315	US-08-793-035-10	Sequence 10, Appl
17	226.5	17.9	244	US-08-375-962B-13	Sequence 13, Appl
18	226.5	17.9	244	US-08-562-114B-13	Sequence 13, Appl
19	226.5	17.9	244	US-08-729-594A-13	Sequence 13, Appl
20	226.5	17.9	244	US-08-937-993-13	Sequence 13, Appl
21	224	17.7	261	US-09-468-738A-29	Sequence 29, Appl
22	224	17.7	261	US-09-940-019-29	Sequence 29, Appl
23	223.5	17.6	337	US-08-440-856A-3	Sequence 3, Appl
24	222.5	17.5	257	US-09-134-001C-3562	Sequence 3562, Ap
25	221.5	17.1	252	US-08-822-322-8	Sequence 8, Appl
26	216	17.1	252	US-09-466-109-8	Sequence 8, Appl
27	216	17.1	252	US-09-466-109-8	Sequence 8, Appl

28	214	16.9	256	US-08-594-808B-7	Sequence 7, Appl
29	209.5	16.6	263	US-09-134-001C-3505	Sequence 3505, Ap
30	208.5	16.5	243	US-09-239-052-2	Sequence 2, Appl
31	203	16.0	271	US-07-637-865-2	Sequence 2, Appl
32	202	16.0	256	US-09-504-358-14	Sequence 14, Appl
33	202	16.0	256	US-09-954-314-14	Sequence 14, Appl
34	198.5	15.7	313	US-09-413-814-9	Sequence 9, Appl
35	197.5	15.6	292	US-09-468-738A-2	Sequence 2, Appl
36	197.5	15.6	292	US-09-940-019-2	Sequence 2, Appl
37	197.5	15.6	296	US-09-468-738A-23	Sequence 23, Appl
38	197.5	15.6	296	US-09-940-019-23	Sequence 23, Appl
39	194.5	15.4	255	US-08-815-225-4	Sequence 4, Appl
40	190	15.0	186	US-08-858-207A-270	Sequence 270, App
41	189.5	15.0	283	US-09-134-001C-5346	Sequence 5346, Ap
42	189	14.9	301	US-09-288-143-98	Sequence 98, Ap
43	189	14.9	327	US-08-375-962B-12	Sequence 12, Appl
44	189	14.9	327	US-08-562-114B-12	Sequence 12, Appl
45	189	14.9	327	US-08-729-594A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-385-028-11
; Sequence 11, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwame A. Aldoo
; APPLICANT: Ashish S. Pardekar
; TITLE OF INVENTION: DNA sequence encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jenner Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER TYPE: Floppy disk
; MEDIUM TYPE: PC-DOS/MS-DOS
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-385-028-11
Query Match 28.9%; Score 366; DB 4; Length 248;
Best Local Similarity 37.3%; Pred. No. 3.3e-33;

Matches 91: Conservative 41: Mismatches 104: Indels 8: Gaps 3:

QY 9 MNPPLNGKVAITGAASGIGLOCAKTLIDAGAKVLLIDREGDKLHKVAEL---GENAYA 65
1 MPSALQGVKVALITGASSGIGATFARALAECAVAIAARVREKIRALGDLTLAGAKVHV 60

QY 66 LQIDLFFNNQOVNMLADIIIELAGGLDIFPHNAGAVIGPVAEGSPDWDRLNINAAF 125
61 LEDVDARQGVDAVAATVEALGDLIVNAGIMLGPVEDADDTWTMRMIDNLLGLM 120

QY 126 RCYAVLPHMIIAORSGDIIFTSIAGVVPVIMEPIYTASKFAVOAFVHTTRQYQXGVR 185
121 YMRALPPLHRSK-GTVVGMSIAGVNVNRNNAVYQATKFGVAASETLRQETRGVGR 179

QY 186 VGAVLPEPVYAL---LDMPPAKMEELANGSLMOPLEVASVLEFMVTRSKNVYRDL 241
180 VVVEPPTDTTELGHITHTATKEMEYORISQIRKLOADIAEAVRYAVAPHHATVHEI 239

QY 242 VILP 245
240 FIRP 243

RESULT 2
US-09-134-001C-4397
Sequence 4397 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: CTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4397
LENGTH: 231
TYPE: PRN
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4397

Query Match 25.3%; Score 319.5; DB 4; Length 231;
Best Local Similarity 30.2%; Pred. No. 5.3e-28;
Matches 70; Conservative 59; Mismatches 92; Indels 11; Gaps 2;

QY 16 KVAITGAASGIGLOCAKTLIDAGAKVLLIDREGDKLHKVAELGENAYALQIDLFFNNQO 75
8 KVAAYVTGASSGIGALVANKLSQOGASIVLVGRNEQRLENYOQLNNDPAKVYTAADVTKSN 67

QY 76 VDNMLADIIELAGGLDIFPHNAGAVIGPVAEGSPDWDRLNINAAFRCYAVLPHM 135
68 IDMLKAVIDHFGIIDLIVNASAGOSLSKTDIVNEQMDIMDIVNIGTLHVLQATLPYL 127

QY 136 IAOISGDIIFTSIAGVVPVIMEPIYTASKFAVOAFVHTTRQYQXGVRVAVLPGPVY 195
128 LKQSSGHIIINLASVSGEPRTKTNNAVYCATKAIAHITQSLKLEKALPTGKVTSTSPQWVD 187

QY 196 TALLD--DMKAKMEELANGSLMOPLEVASVLEFMVTRSKNVYRDLVILP 245
188 TPMTGETDGFERRKLEAO-----NIDAVVAYALTOPSHVNVNVEVTIRP 230

FILE REFERENCE: P2002PL
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11

1	EARLIER APPLICATION NUMBER: 60/043,671
2	EARLIER FILING DATE: 1997-04-11
3	EARLIER APPLICATION NUMBER: 60/043,674
4	EARLIER FILING DATE: 1997-04-11
5	EARLIER APPLICATION NUMBER: 60/043,669
6	EARLIER FILING DATE: 1997-04-11
7	EARLIER APPLICATION NUMBER: 60/043,312
8	EARLIER FILING DATE: 1997-04-11
9	EARLIER APPLICATION NUMBER: 60/043,313
10	EARLIER FILING DATE: 1997-04-11
11	EARLIER APPLICATION NUMBER: 60/043,672
12	EARLIER FILING DATE: 1997-04-11
13	EARLIER APPLICATION NUMBER: 60/043,315
14	EARLIER FILING DATE: 1997-04-11
15	EARLIER APPLICATION NUMBER: 60/048,974
16	EARLIER FILING DATE: 1997-06-06
17	EARLIER APPLICATION NUMBER: 60/056,886
18	EARLIER FILING DATE: 1997-08-22
19	EARLIER APPLICATION NUMBER: 60/056,877
20	EARLIER FILING DATE: 1997-08-22
21	EARLIER APPLICATION NUMBER: 60/056,889
22	EARLIER FILING DATE: 1997-08-22
23	EARLIER APPLICATION NUMBER: 60/056,893
24	EARLIER FILING DATE: 1997-08-22
25	EARLIER APPLICATION NUMBER: 60/056,630
26	EARLIER FILING DATE: 1997-08-22
27	EARLIER APPLICATION NUMBER: 60/056,878
28	EARLIER FILING DATE: 1997-08-22
29	EARLIER APPLICATION NUMBER: 60/056,662
30	EARLIER FILING DATE: 1997-08-22
31	EARLIER APPLICATION NUMBER: 60/056,872
32	EARLIER FILING DATE: 1997-08-22
33	EARLIER APPLICATION NUMBER: 60/056,882
34	EARLIER FILING DATE: 1997-08-22
35	EARLIER APPLICATION NUMBER: 60/056,637
36	EARLIER FILING DATE: 1997-08-22
37	EARLIER APPLICATION NUMBER: 60/056,903
38	EARLIER FILING DATE: 1997-08-22
39	EARLIER APPLICATION NUMBER: 60/056,888
40	EARLIER FILING DATE: 1997-08-22
41	EARLIER APPLICATION NUMBER: 60/056,879
42	EARLIER FILING DATE: 1997-08-22
43	EARLIER APPLICATION NUMBER: 60/056,880
44	EARLIER FILING DATE: 1997-08-22
45	EARLIER APPLICATION NUMBER: 60/056,894
46	EARLIER FILING DATE: 1997-08-22
47	EARLIER APPLICATION NUMBER: 60/056,911
48	EARLIER FILING DATE: 1997-08-22
49	EARLIER APPLICATION NUMBER: 60/056,636
50	EARLIER FILING DATE: 1997-08-22
51	EARLIER APPLICATION NUMBER: 60/056,874
52	EARLIER FILING DATE: 1997-08-22
53	EARLIER APPLICATION NUMBER: 60/056,910
54	EARLIER FILING DATE: 1997-08-22
55	EARLIER APPLICATION NUMBER: 60/056,864
56	EARLIER FILING DATE: 1997-08-22
57	EARLIER APPLICATION NUMBER: 60/056,631
58	EARLIER FILING DATE: 1997-08-22
59	EARLIER APPLICATION NUMBER: 60/056,845
60	EARLIER FILING DATE: 1997-08-22
61	EARLIER APPLICATION NUMBER: 60/056,892
62	EARLIER FILING DATE: 1997-08-22
63	EARLIER APPLICATION NUMBER: 60/057,761
64	EARLIER FILING DATE: 1997-08-22
65	EARLIER APPLICATION NUMBER: 60/047,595
66	EARLIER FILING DATE: 1997-05-23
67	EARLIER APPLICATION NUMBER: 60/047,599
68	EARLIER FILING DATE: 1997-05-23
69	EARLIER APPLICATION NUMBER: 60/047,588
70	EARLIER FILING DATE: 1997-05-23
71	EARLIER APPLICATION NUMBER: 60/047,585
72	EARLIER FILING DATE: 1997-05-23
73	EARLIER APPLICATION NUMBER: 60/047,586
74	EARLIER FILING DATE: 1997-05-23

EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,590
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,594
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,589
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,593
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,614
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,578
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/043,576
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/047,501
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,670
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/056,632
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,664
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,876
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,881
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,909
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,875
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,862
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,887
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,908
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/048,964
EARLIER	FILING DATE:	1997-06-06
EARLIER	APPLICATION NUMBER:	60/057,650
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/056,884
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/057,669
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/049,610
EARLIER	FILING DATE:	1997-06-13
EARLIER	APPLICATION NUMBER:	60/061,060
EARLIER	FILING DATE:	1997-10-02

Query Match	24.4%;	Score 309;	DB 4;	Length 350;
Best Local Similarity	33.7%;	Pred. No. 1.5e-26;		
Matches	83;	Conservative 38;	Mismatches 107;	Indels 18; Gaps 5

Qy	13	LNGLVVAALITGGAASIGTQCARTLDDAARVVLIDREDRLHRLVAEL-----GENA	63
Db	35	LRNVVVITGTATSGTIGKECAKVFYAAAKLVLGCRNGALEELLRELTASHATKRVQTHKP	94
Qy	64	YALDILFNNOQVNMJLADIIETLGGDILFIHANNGAYIGGFVAEGDPVDVRLNLNLINA	123
Db	95	YLVFEDLTDSGAIVAAAAEIIQCIFYVDIILVNNNGISVYRGTIMDTVDVDRKVEVETNYFG	154
Qy	124	AFCRVRAVLPHIMINORSGDIIIFTSINGVAVVPIEPIYTSKFAVQFVHTTROVSQYG	183
Db	155	PVALTKALPLSMIKRQGHIVAISSIGQKMSIPFRSAVYASKNHTQAFPCCLRAEMQYE	214
Qy	184	VRVCAGVLGPV-----VTLALDMPK-AKKEEALANGSLMOPIVAAASYLFMV-TRSKNV	236
Db	215	IEVIVISGYIHTNLVSNALITADSSRYGVMDTTAQC--RSPEVEADVDILAANGKKKKDV	272
Qy	237	TVRDV 242	
Db	273	ILADLL 278	

Db 63 ADMSDGE---ACRALIETAGCDILVNMAGIOHVSSIEEPVVKWMAILINLSAFHT 118
Qy 128 VRAVLPHMIAORSGLIETFTSSAGVAVPIWEPIYASKFAVAQAFVHTTRVSOQYGVAG 187
Db 119 TAAALPGMRKAGMRIVINIASAGLITSPYKSAVAAKHGVGTTKYTALETACKGLTCN 178
Qy 188 AVLPGPVVTALLD 200
Db 179 AICPGVLTPLVE 191

RESULT 11

US-08-440-856A-4
Sequence 4, Application US/08440856A
Patent No. 5750873
GENERAL INFORMATION:
APPLICANT: DELLAPORTA, STEPHEN L. METHODS FOR PRODUCING
TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE. N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,856A
FILING DATE: 15-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 05463-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1517
TELEFAX: (202) 887-0763
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-440-856A-4

Query Match 19.4%; Score 245.5; DB 1; Length 333;

Best Local Similarity 27.5%; Pred. No. 2.1e-19;
Matches 77; Conservative 46; Mismatches 117; Indels 40; Gaps 6;

Qy 3 NHSVPSMNTPLNGKVAATGAASGIGLQCAKTLTDAGAKVLLIDREGDKLHKIVAEIGEN 62
Db 42 NGAPTPMPKRGKVAIVTGGARGIGAIYVLFVKHGAKVVIADIDDAAGFALAALGPH 101
Qy 63 AYALQDLDFNNQVDNMLADIIELAGGLDIFHANAGV-----IGPYAEDDPVMDRVL 117
Db 102 VGFRCVSVVEEDVERAVERAVARYGRDLVLCNNAGVLGROTRAASKITLSPDAGEPDRVL 161
Qy 118 NLN-INAFAFCVRAVLPHMIAORSGLIETFTSSAGVAVPIWEPIYASKFAVAQAFVHTTR 176
Db 162 RVNALGAAALGKHAALA-MTORRAGSIISVAVAGVLGGLGPHAYTASKAHIVGLTKNAA 220
Qy 177 ROVSQYGVRAVLPGPVVTALL-----DDMPKAKME 209
Db 221 CEIGAHGIRVNCISPFVAVAPMLINAWRGHDASTADADADIDLDAVPSDQVEKME 280

Qy 210 -----ALANGSLMOPIEVAESVLEMTV-RSKNVTVRDLVI 243
Db 281 VVRGLATLKGAATLRPRDIAEALFLASDDSRYSIGHNLVV 320

RESULT 12

US-09-134-001C-4431
Sequence 4431, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4431
LENGTH: 274
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4431

Query Match 19.2%; Score 242.5; DB 4; Length 274;

Best Local Similarity 28.3%; Pred. No. 3.4e-19;
Matches 71; Conservative 50; Mismatches 113; Indels 17; Gaps 5;

Qy 13 LNKGVAAITGASGIGLQCAKTLTDAGAKVLLIDREGDKLHKIVAEIGEN--AYALQDL 69
Db 6 LENKIAVITGASTGIGASAVALAIEGAHVLAID-ISDLEETVQSIINDNGKATAVRVD 64
Qy 70 LFNNOQVDNMLADIIELAGGLDIFHANAGVIG-CPYAEDGDPVMDRVNLINIAAFRCV 128
Db 65 ISDDKQVKPSEKTAQERGHADVFNNAVDNGAGRITHEPVEFDKIMAVDMKGTFLVT 124
Qy 129 RAVLPHMIAORSGLIETFTSSAGVAVPIWEPIYASKFAVAQAFVHTTRVSOQYGVAG 188
Db 125 KFLPLMKQ-KGSIINTASGGAADLYSGVAAKGVINTFKSIAIEYGRINIRANA 183
Qy 189 VLPGPVVTALLDMPKAKMEBA-----LANGSLMOPIEVAESVLEMTVRSKNVT 237
Db 184 IAPGTIETPLVDNLACTSDEAGQTFRENQKWTPLGRLGTPDEVGLVAFILASDDSSFI 243
Qy 238 VRDLVILPGSV 248
Db 244 TGETIRIDGV 254

RESULT 13

US-09-363-189B-6
Sequence 6, Application US/09363189B
Patent No. 6242228
GENERAL INFORMATION:
APPLICANT: SUGIYAMA, MASAKAZU
APPLICANT: TONOUCHI, NAOTO
APPLICANT: SUZUKI, SHUNICHI
APPLICANT: YOKOZEKI, KENZO
TITLE OF INVENTION: XLITROL DEHYDROGENASE OF ACETIC ACID BACTERIA AND GENE THERE
FILE REFERENCE: 0010-1024-0
CURRENT APPLICATION NUMBER: US/09/363,189B
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: JP10-216047
PRIOR FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 262
TYPE: PRT
ORGANISM: Glucobacter oxydans

US-09-363-189B-6

Query Match 19.1%; Score 242; DB 4; Length 262;
 Best Local Similarity 34.4%; Pred. No. 3.6e-19;
 Matches 65; Conservative 28; Mismatches 90; Indels 6; Gaps 3;

QY 9 MNTPLNGVVAITGAASGIGLOCAKTLTDAGAKVVLIDREGDKLHKI--VAELGENAYA 65
 DB 1 MSKRFNGKVLCTGAGNGNIGLATALRLAEGETAIALLDNMNRALKAEKAEVREKYEARS 60
 QY 66 LQDLFNNQVDNMLADIIELAGGLDIFHANAGAYIG--GPAEEDPDVMDVNLININA 123
 DB 61 YVCDVTSEAVIGTVDSVVRDCKIDFLFNAG--YOGAFAPQDPSDDFAVRLTINVTG 119
 QY 124 AFRCRAVLPFHIAORSGDIIFTSSIAGVVPVIWEPIYASKFAVOAFVHTTRROVSOGX 183
 DB 120 AFVYKAASRQMTITONIGRIYATWASAGYKPPNMAAYGASKGAILTETALDIALPI 179
 QY 184 VRVGAVALPG 192
 DB 180 IRVNAISPG 188

RESULT 14
 US-09-134-001C-5042

; Sequence 5042, Application US/09134001C
 ; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5042

; LENGTH: 267

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5042

Query Match 19.1%; Score 241; DB 4; Length 267;
 Best Local Similarity 30.6%; Pred. No. 4.8e-19;
 Matches 72; Conservative 39; Mismatches 100; Indels 24; Gaps 7;

QY 13 LNKVAALITGAASGIGLOCAKTLTDAGAKVVL---IDREGDKLHKI--VAELGENAYA 66
 DB 9 LENKVVLTGATGTGIGKSTAEENFGAKAKAVVINYSDRHSSELEKQYVAFGGCTTAV 68
 QY 67 QDLFNNQVDNMLADIIELAGGLDIFHANAGAYIGGPAEEDPDVMDVNLININA 126
 DB 69 QODVASEIEDIKRIETFTIHFGFLDIINNAGFENSIPTHEMSIDDMQVITDINLGAVY 128
 QY 127 CVRAVLPFHIAO--RSGDIIFTSSIAGVVPVIWEPI--YASKFAVOAFVHTTRROVSOGX 183
 DB 129 GSEETINFLKRNKGTIINISSVHTIP--WPNYVHYAASKGKLIMETSMSEYAOG 186
 QY 184 VRVGAVALPGVVTALDDWPKAMEBALANGSLMOP-----EVAESVLFM 229
 DB 187 IRINNISPGAIYV---EHTKEFSDPTTREETIKMIPAREIGNADVANAVALFL 237

RESULT 15
 US-08-793-035-9

; Sequence 9, Application US/08793035

; Patent No. 6011201

; GENERAL INFORMATION:

; APPLICANT: Stabas, Antoni R.

; APPLICANT: White, Andrew

; APPLICANT: Chase, Dianne
 ; APPLICANT: Elborough, Keiran
 ; APPLICANT: Fentem, Philip A.
 ; TITLE OF INVENTION: B-ketacyl ACP Reductase Genes From
 ; TITLE OF INVENTION: Brassica Napus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: TX

COUNTRY: US
 ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/793,035

FILING DATE: 28-JUL-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9414622.2

FILING DATE: 20-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB95/01678

FILING DATE: 17-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kammerer, Patricia A.

REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: MOBT.132

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713.787.1400

TELEFAX: 713.787.1440

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 315 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-793-035-9

Query Match 18.3%; Score 231.5; DB 3; Length 315;
 Best Local Similarity 28.9%; Pred. No. 7.2e-18;
 Matches 67; Conservative 43; Mismatches 111; Indels 11; Gaps 3;

QY 5 SVPSMNTPLNGVVAITGAASGIGLOCAKTLTDAGAKVVL---IDREGDKLHKIYAEIG 60
 DB 66 AVPKVESP---VVVYVTGASRGIGKAIALSLKAGCKVLVYNARSAKKEAEVSKQIEAVG 121
 QY 61 ENAYALQDLFNNQVDNMLADIIELAGGLDIFHANAGAYIGGPAEEDPDVMDVNLIN 120
 DB 122 GOAITFGGCVSKEADEAMKMTAIDAMGTIDVYVNNAGITRDTLLRMKKSOMDEVIDLN 181
 QY 121 INAAFCRAVLPFHIAORSGDIIFTSSIAGVVPVIWEPIYASKFAVOAFVHTTRROVS 180
 DB 182 LTGVFLCTQDAATKIMKKKRRGRININIASVGLGIGNANTAAKAGVIGFSKTAAREGA 241
 QY 181 QXGVRAVLPGPV---VTALDDWPKAMEBALANGSLMOPLEVAESVLFM 229
 DB 242 SRINNNVVCPCFIASDMTAKIGEDMEKKILGTIPLGRYGQPEDVAGLVEFL 293

Search completed: March 13, 2003, 16:42:10
 Job time: 7.82947 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 7.03642 Seconds
(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHSHVSPSMNTPLNCKVAAL.....TRSKNVTVRLVILPGSYDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCIT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCITUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	9	US-09-802-208B-3
2	311	24.6	337	9	US-09-860-670-143
3	311	24.6	337	10	US-09-764-853-498
4	309	24.4	310	9	US-09-905-291A-153
5	309	24.4	310	9	US-09-902-853-153
6	309	24.4	310	9	US-09-907-824-153
7	309	24.4	310	9	US-09-907-841-153
8	309	24.4	310	9	US-09-904-011-153
9	309	24.4	310	9	US-09-906-742-153
10	309	24.4	310	9	US-09-906-838-153
11	309	24.4	310	9	US-09-907-613-153
12	309	24.4	310	9	US-09-907-942-153
13	309	24.4	310	9	US-09-904-820-153
14	309	24.4	310	9	US-09-904-859-153
15	309	24.4	310	9	US-09-909-204-153
16	309	24.4	310	9	US-09-904-786-153
17	309	24.4	310	9	US-09-906-646-153
18	309	24.4	310	9	US-09-906-700-153
19	309	24.4	310	9	US-09-902-903-153

20	309	24.4	310	9	US-09-903-749A-153	Sequence 153, App
21	309	24.4	310	9	US-09-903-786-153	Sequence 153, App
22	309	24.4	310	10	US-09-909-320-153	Sequence 153, App
23	309	24.4	310	10	US-09-909-088B-153	Sequence 153, App
24	291.5	23.0	254	9	US-09-978-758-2	Sequence 2, Appl1
25	281.5	22.3	203	10	US-09-823-901-7	Sequence 7, Appl1
26	278	22.0	206	9	US-09-997-816-4	Sequence 4, Appl1
27	270.5	21.4	251	10	US-09-815-242-5689	Sequence 5689, Ap
28	270.5	21.4	272	10	US-09-815-242-12688	Sequence 12688, A
29	255	20.2	246	10	US-09-815-242-5461	Sequence 5461, Ap
30	255	20.2	246	10	US-09-815-242-12123	Sequence 12123, A
31	255	20.2	246	10	US-09-815-242-12803	Sequence 12803, A
32	255	20.2	246	10	US-09-815-242-13100	Sequence 13100, A
33	251	19.8	263	10	US-09-815-242-14092	Sequence 14092, A
34	250.5	19.8	266	9	US-09-981-353-40	Sequence 40, Appl1
35	250.5	19.8	266	10	US-09-931-186-18	Sequence 18, Appl1
36	250.5	19.8	276	9	US-09-944-160-24	Sequence 24, Appl1
37	249.5	19.7	266	9	US-09-885-303A-22	Sequence 22, Appl1
38	249	19.7	261	10	US-09-815-242-5813	Sequence 5813, Ap
39	247	19.5	258	10	US-09-954-314-12	Sequence 12, Appl1
40	245	19.4	253	10	US-09-815-242-11842	Sequence 11842, A
41	242	19.1	262	10	US-09-802-853-6	Sequence 6, Appl1
42	242	19.1	312	9	US-09-885-303A-16	Sequence 16, Appl1
43	241.5	19.1	278	9	US-10-033-245-2	Sequence 2, Appl1
44	241.5	19.1	278	9	US-10-033-223-2	Sequence 2, Appl1
45	241.5	19.1	278	9	US-10-033-167-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-09-802-208B-3
Sequence 3, Application US/098022208B
Publication No. US20030041352A1
GENERAL INFORMATION:
APPLICANT: Parrott, Wayne
APPLICANT: Lafayette, Peter
APPLICANT: Kane, Patrick
TITLE OF INVENTION: Antibiotol or Ribitol As Positive Selectable Markers
FILE REFERENCE: UGA 855R
CURRENT APPLICATION NUMBER: US/09/802,208B
CURRENT FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 250
TYPE: PRT
ORGANISM: Escherichia coli
US-09-802-208B-3

Query Match	100.0%	Score 1265;	DB 9;	Length 250;
Best Local Similarity	100.0%	Pred. No. 1,2e-117;		
Matches 250;	Conservative	0;	Mismatches 0;	Indels 0;
			Gaps	0;
OY	1	MNHSHVSPSMNTPLNCKVAALITGAASIGLOCAKTLIDAGAKVVLIDREGDKLHKIYAELG	60	
Db	1	MNHSHVSPSMNTPLNCKVAALITGAASIGLOCAKTLIDAGAKVVLIDREGDKLHKIYAELG	60	
OY	61	ENAYALQDLFPNQOVNMLADIELAGGLDIFHANAGAYIGGPAVEGDPDWDVRLNIN	120	
Db	61	ENAYALQDLFPNQOVNMLADIELAGGLDIFHANAGAYIGGPAVEGDPDWDVRLNIN	120	
OY	121	INAAFCVRAVLPHMTAORSSGDIIFTSISAGVVPVYTWEPYITASKRAVAFVHTTRRVS	180	
Db	121	INAAFCVRAVLPHMTAORSSGDIIFTSISAGVVPVYTWEPYITASKRAVAFVHTTRRVS	180	
OY	181	QGVGVGAVLPQGVYVYALDDMPKAKMEBALANGSLMOPIEVAESVLFMTTRSKNTVVD	240	
Db	181	QGVGVGAVLPQGVYVYALDDMPKAKMEBALANGSLMOPIEVAESVLFMTTRSKNTVVD	240	
OY	241	VILPGSYDL 250		


```

APPLICANT: Tumias, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,853
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US/09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-902-853-153

Query Match      24.4% Score 309; DB 9; Length 310;
Best Local Similarity 33.7%; Pred. No. 7.4e-23;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

13  LNKGAATATGAASGIGLCACAKTLIDAGAKVVLIDREGDKLHIVAEI-----GENA 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35  LKNAYVVTIGATSGIGKCKCAKVFYAAGKIVILCGNGGALBELIELTASHATKYQTHKP 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64  YALQIDLFENNOQVDMMLADIIEFLAGGLDIFHANAGAYIGGVAESDPDVMRYLNLNTNA 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95  YLVTFDLIDSGAIVAAAEIILQCFGYVDILVNNAGISYRGITMDTVVDVKRMETNFG 154
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124  AFRCRAVLAIPHMLAORSDDIIFTSSIAGVPYIMPEITYASKFAVOAVVHTTRQVSOG 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155  PYALTKLRKSKIKRQGHIVAISSIQGKMSIIPFSAYTAASHAHATQAFDFDLRAEMOTE 214
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184  VRVGAVLGDPV-----VALLDDMPK-AKMEBALANGSLIMPIEVAESYLEWV-TRSKNV 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215  IEVTVISPGYIHTLSVNAITADGSRYGVMOTTTAG--RSPVEVAQDVLAAGKKKRDV 272
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 237  TVRDLY 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273  ILADLL 278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6
US-09-907-824-153
Sequence 153, Application US/09907824
Publication No. US20020197671A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathier, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,824
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423

SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-824-153
Query Match 24.4%; Score 309; DB 9; Length 310;
Best Local Similarity 33.7%; Pred. No. 7,4e-23;
Matches 83; Conservative 36; Mismatches 107; Indels 18; Gaps 5;
-GENA 63
13 LNKVAAITGASGIGCACTLLDAGAKVYLIDREDPKLHKIYAEI-----
Db 35 LRNAVVTTCATSGLGECAKVFYAAGAKVLGCRNGALBELRELTAHSATRVQTHKP 94
Qy 64 YALDLFNNQVDNMLADIIETLAGDIFRANAGAYIGFVAGDDPDVMDRVNLNMA 123
Db 95 YLVTFDLDSGAIYAAAEIETGCGYVDLVNNGISRGITMDTVDOVKRWETVFG 154
Qy 124 AFRCRAVLPHMIAQRSGDIIETSSINAGVVPVIMEPIYTSKFAVQAFVHTTRQVSQYG 183
Db 155 PVALTAKLLPSMIKRRGHIYAIISSIOGKMSIPRSAYVAASKHATQAFDFCLRAEMEOYE 214
Qy 184 VRGAVLPGPY-----VTALDDMPK-AKMEALANGSLMQPIEVAESVLFMV-TRSKNV 236
Db 215 LEVTVISPGYHTNLVSNALTAAGSRGVMDTTTAAQ--RSPVEADVDLAAGKKKKDV 272
Qy 237 TVRDLV 242
Db 273 ILADLL 278
RESULT 7
US-09-907-841-153
Sequence 153, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathier, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222

APPLICANT:	Genentech, Inc.
APPLICANT:	Ashkenazi, Avi
APPLICANT:	Botstein, David
APPLICANT:	Desnoyers, Luc
APPLICANT:	Falon, Dan L.
APPLICANT:	Ferrira, Napoleone
APPLICANT:	Filvaroff, Ellen
APPLICANT:	Fong, Sherman
APPLICANT:	Gao, Wei-Qiang
APPLICANT:	Gerber, Hanspeter
APPLICANT:	Gertlisen, Mary E.
APPLICANT:	Goddard, A.
APPLICANT:	Godowski, Paul J.
APPLICANT:	Grimaldi, Christopher J.
APPLICANT:	Gurney, Austin L.
APPLICANT:	Hillan, Kenneth, J.
APPLICANT:	Kijavlin, Ivar J.
APPLICANT:	Mather, Jennie P.
APPLICANT:	Pan, James
APPLICANT:	Pavoni, Nicholas F.
APPLICANT:	Roy, Margaret Ann
APPLICANT:	Stewart, Timothy A.

```

APPLICANT: Williams, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
FILE OF INVENTION: Acids Encoding the Same
CURRENT APPLICATION NUMBER: US/09/904,011
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-011-153

Query Match      24.4%; Score 309; Length 310;
Best Local Similarity   33.7%; Pred. No. 7,4e-23;
Matches    83; Conservative    38; Mismatches    107; Indels    18; Gaps    5

OY      13 LNKVAAITGAASGIGLOCAKTLLDAGAKVLLIDREGDLRHIVAEI-----GENA 63
           ||| | | | | | | | | | | | | | | | : : : |
DB      35 LRNAVVTIGATSGICKGEKAFVAAGAIVLCLGRNGCALLEELIRELTASHATKYOTHRP 94
           :||| |:| | | | | | | | | | | | | | | | : : : |
OY      64 YALQJLDENNOCVDMMLADIIELAGGLDIFHANAGYIGGPAAEGDPDWDRVLNINIA 123
           | : | : | : | : | : | : | : | | | | | | : | | : |
DB      95 YLVFTDLDSGIVAAMAEILLCFGYVDILLVNNAGISYRGTTMDTVDVDDKRVMEIFYG 154
           | : | : | : | : | : | : | : | | | | | | | | | |
OY      124 AFRCRYAVLPPIHIAQRSGDIIFTSTIAGVVPVIWEPIYASKFAVOAFHTTRROYSOYG 183
           | : | : | : | : | : | : | : | | | | | | | | | |
DB      155 PVALTALLPSHIKRQGIIIVAISSIQGMSPFRRSAIVASKAHATDAFFDCRLRAEMEOTE 214
           | : | : | : | : | : | : | : | | | | | | | | | |
OY      184 VRVGVALPGPV-----VTALLDDWPK-AKMEBALANGSLMOPLEVASVLFMY-TRSKNV 236
           | : | : | : | : | : | : | : | | | | | | | | | |
DB      215 IEVTVISGYIITNLNVNAITADGSRYGVMDTTTAQG--RSPVEVAQDVLAAGKKKKDY 272
           | : | : | : | : | : | : | : | | | | | | | | | |
OY      237 TVRDLY 242 ,
           : || :

```


Sequence 155, Application US/099507613
Publication No. US20030027145A1
GENERAL INFORMATION:
APPLICANT: Gentech, Inc.
APPLICANT: Ashtkenazi, Avi
APPLICANT: Boststein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.

Qy 13 LNKGAATTTGAAGSGIGLCCAKTLLDAGAKVVLIDRECDKLHKIYAE-----GENA 63

Db 35 LRNAVVITGATSGLGKRECAKVFYAAGAKVLVCGRNGALEELIRELTASHATKVQTHKP 94

	PRIOR APPLICATION NUMBER:	PCT/US99/28313	
	PRIOR FILING DATE:	1999-11-30	
	PRIOR APPLICATION NUMBER:	PCT/US99/28564	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/28565	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/30095	
	PRIOR FILING DATE:	1999-12-16	
	PRIOR APPLICATION NUMBER:	PCT/US99/30911	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US99/30999	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US00/00219	
	PRIOR FILING DATE:	2000-01-05	
	NUMBER OF SEQ ID NOS:	423	
	SEQ ID NO:	153	
	LENGTH:	310	
	TYPE:	PRT	
	ORGANISM:	Homo sapiens	
	DB:	US-09-907-942-153	
	Query Match	24.4% Score 309; DB 9; Length 310;	
	Best Local Similarity	33.7% Pred. No. 7.4e-23;	
	Matches	83; Conservative 38; Mismatches 107; Indels 18; Gaps	
OY	13	LNGRYAAITGAASGIGLOCAKTLIDAGAKVVLIDREGBLHKIVAEI-----GENA	63
Dd	35	LNNAVVITGATSGIGKECAKFYAAQAKLVLCGRNNGALELLIRELPASHATKVOTHRP	94
OY	64	VALDLDLENNQOVDMKLDIETLAGGLDIFHNAGAYIGGPVAEGDPVDVNLINITA	123
Dd	95	YLVTDLTDSGAIYAAAEHLIOCFEYVDILVNNAGISIRGTIMDTTVDDKKRMETTNFG	154
OY	124	AFCRYAVLPHMIAORSODIIFTSSIAGVPVIWEPIYTASKFAVOAFVHTTRROVSOGX	183
Dd	155	PVALTKALLPSMKRRQGHIYAISSIQCKMSIPFRSAVAASHKHATQAFEDCLRAEMGEY	214
OY	184	VRYGVAVLEGPV-----VALLDPMRK-AKMEALANGSLMOPIEVAESVLFMV-TRSKNV	236
Dd	215	LEVTVISGCIHNTLSVNAITADSGSRYGVMITTQAG--RSPVEVAQDVLAVGKRKKDY	272
OY	237	TVRDLV	242
Dd	273	ILADDL	278
 RESULT 13 US-09-904-820-153 Sequence 153, Application US/09904820 Publication No. US20030036094A1 GENERAL INFORMATION: APPLICANT: Genentech, Inc. APPLICANT: Ashkenazi, Avi APPLICANT: Botstein, David APPLICANT: Desnoyers, Luc APPLICANT: Eaton, Dan L. APPLICANT: Ferrara, Napoleone APPLICANT: Filvaroff, Ellen APPLICANT: Fong, Sherman APPLICANT: Gao, Wei-Qiang APPLICANT: Gerber, Hanspeter APPLICANT: Gerritsen, Mary E. APPLICANT: Goddard, A. APPLICANT: Godowski, Paul J. APPLICANT: Grimaldi, Christopher J. APPLICANT: Gurney, Austin L. APPLICANT: Hillan, Kenneth, J. APPLICANT: Kljavin, Ivar J. APPLICANT: Mather, Jennie P. APPLICANT: Pan, James APPLICANT: Paoni, Nicholas F. APPLICANT: Roy, Margaret Ann APPLICANT: Stewart, Timothy A.			

APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,820
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-820-153

Query Match 24.4%; Score 309; DB 9; Length 310;
Best Local Similarity 33.7%; Pred. No. 7.4e-23;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

13 LNKVAATGAAGSGIGLQCAKTLIDREGDKLHKIVAEI-----GENA 63
13 LNKVAATGAAGSGIGLQCAKTLIDREGDKLHKIVAEI-----GENA 63
35 LRNAVVITGATSGIGLQCAKTLIDREGDKLHKIVAEI-----GENA 94
35 LRNAVVITGATSGIGLQCAKTLIDREGDKLHKIVAEI-----GENA 94
64 YALOLDLNNQOVNMLADIIELAGLIDFHNAGAYIGPVAEGDPDWDVRLNINA 133
64 YALOLDLNNQOVNMLADIIELAGLIDFHNAGAYIGPVAEGDPDWDVRLNINA 133
95 YLTFPDLTSGAIVAAAEELIQCFGYVDILVNNAGISYRGTIMDTVDVKRVMETNYFG 154
95 YLTFPDLTSGAIVAAAEELIQCFGYVDILVNNAGISYRGTIMDTVDVKRVMETNYFG 154
124 AFGCAVAVLPHMTAORSGLIIFTSIAGVYVIMPEIYTSKRVAVQAFVHTTRVSOYG 163
124 AFGCAVAVLPHMTAORSGLIIFTSIAGVYVIMPEIYTSKRVAVQAFVHTTRVSOYG 163
155 PVALTFALLPSTMRKROGHIVAISSIOGKMSIPRSAYVASKHATQAFDCLRAEMOYE 214
155 PVALTFALLPSTMRKROGHIVAISSIOGKMSIPRSAYVASKHATQAFDCLRAEMOYE 214
184 VRGCAVLPGFV-----VTALLDMPK-AKMEALANGSLQPIEVAESVLFMV-TRSKNV 236
184 VRGCAVLPGFV-----VTALLDMPK-AKMEALANGSLQPIEVAESVLFMV-TRSKNV 236
215 IETVIVSPGIYHTNLSTNATITADGSRGVMDTTAQS--RSPVEVADVYLAAGKKKKKV 272
215 IETVIVSPGIYHTNLSTNATITADGSRGVMDTTAQS--RSPVEVADVYLAAGKKKKKV 272
237 TVRDLV 242
237 TVRDLV 242

DB 273 ILADLL 278
RESULT 14
US-09-904-859-153
Sequence 153, Application US/09904859
Publication No. US20030036060A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,859
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 8.69205 Seconds
(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHNSVPSMNTPLNGKVAAL.....TRSKAVTVRDVLTPGSVDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1100	87.0	249	1 S07134	ribitol 2-dehydrog
2	1073	84.8	247	1 DEKBR	ribitol 2-dehydrog
3	843	66.6	242	2 AC3632	ribitol 2-dehydrog
4	573	45.3	243	2 B98199	ribitol 2-dehydrog
5	573	45.3	243	2 AG3087	ribitol 2-dehydrog
6	333	26.3	249	2 E95283	probable [imported
7	331	26.2	231	2 G90050	hypothetical prote
8	304	24.0	325	2 T17307	hypothetical prote
9	295.5	23.4	255	2 T44991	oxidoreductase [im
10	295	23.3	248	1 B39930	Intermin B - Lis
11	294	23.2	248	2 AD1489	oxidoreductase hom
12	292	23.1	248	2 AD1128	oxidoreductase hom
13	289	22.8	247	2 AG1672	oxidoreductase, sh
14	288.5	22.8	252	2 B82181	3-oxoacyl-(acyl)-ca
15	288	22.8	248	2 T44932	3-oxoacyl-(acyl)-ca
16	287.5	22.7	246	2 C83961	3-oxoacyl-(acyl)-ca
17	286	22.6	247	2 F84210	oxidoreductase hom
18	285	22.5	247	2 AG1300	3-ketoacyl-acyl ca
19	284	22.5	246	2 A69621	3-oxoacyl-(acyl)-ca
20	283	22.4	246	2 H72219	3-oxoacyl-(acyl)-ca
21	282	22.3	284	2 T02257	probable short cha
22	281	22.2	233	2 T35242	probable oxidoredu
23	280	22.1	255	2 AF2466	hypothetical prote
24	279.5	22.1	286	2 AI3207	NAD/NADP dependent
25	278.5	22.0	244	2 S76993	hypothetical prote
26	278	22.0	246	2 H84136	3-oxoacyl-(acyl)-ca
27	278	22.0	247	2 F82643	oxidoreductase xfl
28	278	22.0	255	2 T51764	probable 2-hydroxy
29	277	21.9	246	2 A87457	3-oxoacyl-(acyl)-ca

30	277	21.9	251	2 AH2042	3-oxoacyl-(acyl)-ca
31	275.5	21.8	253	2 B86737	acetoin dehydrogen
32	273.5	21.6	260	2 H70758	probable fabG3 pro
33	272.5	21.5	272	2 A99950	hypothetical prote
34	272	21.5	266	2 F83127	probable short-cha
35	271.5	21.5	260	2 G84171	oxidoreductase [im
36	271	21.4	238	1 D69930	probable 3-oxoacyl
37	270	21.3	249	2 B97605	cyclohexanol dehyd
38	270	21.3	249	2 AD2827	short chain dehydr
39	269	21.3	257	2 AB1866	hypothetical prote
40	268.5	21.2	248	2 F82128	3-oxoacyl-(acyl)-ca
41	268.5	21.2	248	2 F81971	probable 3-oxoacyl
42	268.5	21.2	252	2 H64122	ylfG protein - Hae
43	267	21.1	258	2 H75616	oxidoreductase, sh
44	267	21.1	278	2 E83152	probable short-cha
45	266.5	21.1	244	2 T12051	3-oxoacyl-(acyl)-ca

ALIGNMENTS

RESULT 1
S07134
ribitol 2-dehydrogenase (EC 1.1.1.56) - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 18-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Oct-2000
C:Accession: S07134; S07135
R:Dothie, J.M.; Giglio, J.R.; Moore, C.B.; Taylor, S.S.; Hartley, B.S.
Biochem. J. 230, 569-578, 1985
A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence and properties of w1
A:Reference number: S07134; MUID:86050423; PMID:3904726
A:Accession: S07134
A:Molecule type: protein
A:Residues: 1-249 <DOT>
A:Note: artifactual S-carboxymethyl cysteine was found at positions 30 and 126 after
R:Loviny, T.; Norton, P.M.; Hartley, B.S.
Biochem. J. 230, 579-585, 1985
A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural ge
A:Reference number: S07135; MUID:86050424; PMID:2933028
A:Accession: S07135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145, 'SS', 148-211, 'N', 213-249 <LOW>
A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase
F:15-191/Domain: short-chain alcohol dehydrogenase homology <SADH>
Query Match
Best Local Similarity 87.0%; Score 1100; DB 2; Length 249;
Matches 214; Conservative 21; Mismatches 14; Indels 0; Gaps 0;
OY 2 MNHNSVPSMNTPLNGKVAATGAASGIGLCAKTLLDAGAVVLIIDRGGDLHKTIVAEIGE 61
DB 1 MKHSVSMNTSLSGKVAATGAASGIGLECARLLGLGAVVLIIDRGGELNKLTVAEIGE 60
OY 62 NAVALQDLDFNNQOVNMLADIETELAGGLDIFPANAAGATGCGVPAEGDPVMDRVNLINI 121
DB 61 NAFALQVDLMDAQOVNMLQGLIQLGRDIFPANAAGATGCGVPAEGDPVMDRVNLINI 120
OY 122 NAAFRCRAVLPMLIAQRSGDIIFTSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVSQ 181
DB 121 NAAFRCRAVLPMLIAQRSGDIIFTFAVAGVVPVIMEPIYTASKFAVOAFVHTTRQVAQ 180
OY 182 YGVRGAVLPGRVVTALDDMPRAKMEALANGSLMQPIVAASYLEFMVTRSKNVTVRD 241
DB 181 YGVRGAVLPGRVVTALDDMPRAKMEALANGSLMQPIVAASYLEFMVTRSKNVTVRD 240
OY 242 VILPGSVDL 250
DB 241 VILPNSVDL 249
RESULT 2

DEKBR
ribitol 2-dehydrogenase (EC 1.1.1.56) - Enterobacter aerogenes
C:Species: Enterobacter aerogenes
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 23-Mar-1995
C:Accession: A94585; A00344
R:Hartley, B.S.
submitted to the Atlas, June 1977
A:Reference number: A94585
A:Accession: A94585
A:Molecule type: protein
A:Residues: 1-247 <HAR>
A:Note: this enzyme binds the coenzyme nicotinic adenine dinucleotide
R:Moore, C.H.; Taylor, S.S.; Smith, M.J.; Hartley, B.S.
unpublished results, cited by Morris, H.R., Williams, D.H., Midwinter, G.G., and Hartley,
A:Reference number: A94480
A:Contents: annotation
A:Note: Cys-30 and Cys-126 are carboxymethylcysteine
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase
C:13-185/Domain: short-chain alcohol dehydrogenase homology <SADH>

```

Query Match      84.8%  Score 1073; DB 1; Length 247;
Best Local Similarity 84.3%  Pred. No. 9, 3e-82;
Matches 210; Conservative 23; Mismatches 14; Indels 2; Gaps 1;

OY 2 MNHSPVSNTPPLNGKVAAITGAASGIGLQCAKTLLDAGAKVYLIDREDGKLHKIVAEIGE 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MKHSYNNITSLSGKVAAITGAASGIGLECAKRTLLTGAAKVVYLIDREGEKLNKLVAELGQ 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 62 NAYVALQDLDFNNQVDNNLADIIEELAGGLDIFHNAGAYTGGPVAEISGPDWDVRLNLI 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 NAFALQVDLMQADQVDNLLGSLQLTGRLDIFHNAGAYTGGPVAEISGPDWDVRLNLI 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 122 NAAFCRVAVLPLPHMIAQRSGDIIIFTSSINGVVPVIMEPIYTASKRPAVOAFVHTTRRQVSQ 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 NAAFCRVASVLPPLHLAQKSGDIIIFTAVIAGV--VIMEPVYTASKRPAVOAFVHTTRRQVAQ 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 182 YGVRVGAIVLPBPVVTATLLDDMPKAKMEALANGSLIMOIEVAESVLEFWYTSKNVTADL 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 YGVRVGAIVLPBPVVTATLLDDMPKAKMEALADGSLIMOIEVAESVLEFWYTSKNVTARDI 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 242 VILPGSYDL 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 VILPNSVDL 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AC3632
ribitol 2-dehydrogenase (EC 1.1.1.56) [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AC3632
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
  ; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier,
  Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3632
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-242 <KUR>
A:Cross-references: GB:AE008918; PIDN:AL54222.1; PID:g17985193; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10980
A:Map position: 11
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase

```

[illegible]

```

RESULT 4
B98199
ribitol 2-dehydrogenase (EC:1.1.1.56) [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text-change 11-Jan-2002
C:Accession: B98199
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirallo, B.; Goldm
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: B98199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89116.1; PID:g15158922; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_1076
A:Map position: linear chromosome
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase
Query Match 45.3%; Score 573; DB 2; Length 243;
Best Local Similarity 46.1%; Pred. No. 3,1e-40;
Matches 111; Conservative 52; Mismatches 78; Indels 0; Gaps 0;

```

[illegible]

C:Accession: AG3087
R:Wood, D.W.; Seibald, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG3087
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-243 <KUR>
A:Cross-references: GB:AE008689; PIDN:AL45117.1; PID:g17742787; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: rdh
A:Map position: linear chromosome
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 45.3%; Score 573; DB 2; Length 243;
Best Local Similarity 46.1%; Pred. No. 3.1e-40;
Matches 111; Conservative 52; Mismatches 78; Indels 0; Gaps 0;

Qy 9 MNTPLNGVAATGAAGTGGTCACTLLDAGAKVLLIDREGDKLHKIVAEIGENAVATL 68
Db 1 MSESIGKIVITGAAGTGGTCACTLLDAGAKVLLIDREGDKLHKIVAEIGENAVATL 60
Qy 69 DLFNNQOVNMLADITELAGLDIFHANAGATGGPVAEGDPDVRVLTNLNNAFRCV 128
Db 61 NLLDADSCNAMIPEITIGVVDHIDILYCNAAGTIGDLETTEPEADIKMLNLVNAVMKNV 120
Qy 129 RAVLPFHMAIORSDDITFTSSAGVAVVIMEPTITASKFAVOAFVITTRQVSGVAVRGA 188
Db 121 QAVPVHMEKRTGDIITVCISAGHPPTWEPTVSSKMAITTSFGMRKMPHGVRAQ 180
Qy 189 VLPGPVATLLDDMPKAKMEALANGSLMOPLEVAESVLFVTRSKNTVTRDLVILPGSV 248
Db 181 VSPGVTVALLDMPKAKMEALANGSLMOPLEVAESVLFVTRSKNTVTRDLVILPGSV 240
Qy 249 D 249
Db 241 D 241

RESULT 6
E95283
Probable [Imported] - *Sinorhizobium meliloti* (strain 1021) megaplasmid pSymA
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95283
R:Barnett, M.J.; Fisher, R.H.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
; Kalman, S.; Keating, D.H.; Palm, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: E95283
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-249 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK64831.1; PID:g14523244; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0320

A:Genome: Plasmid
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 26.3%; Score 333; DB 2; Length 249;
Best Local Similarity 34.4%; Pred. No. 2.7e-20;
Matches 84; Conservative 45; Mismatches 105; Indels 10; Gaps 4;

Qy 11 TPLNKGVAATGAAGTGGTCACTLLDAGAKVLLIDREGDKLHKIVAEIGENAVATL 66
Db 2 TSLNKGIVATGAAGTGGTCACTLLDAGAKVLLIDREGDKLHKIVAEIGENAVATL 60
Qy 67 QLDLFNNQOVNMLADITELAGLDIFHANAGATGGPVAEGDPDVRVLTNLNNAFRCV 126
Db 61 EMDVVDITTSVAGVKKLVDAVGSIDILVNNAGLMPISIDIEKVDENQRAVDVAVKGLN 120
Qy 127 CVRAVLPFHMAIORSDDITFTSSAGVAVVIMEPTITASKFAVOAFVITTRQVSGVAVR 185
Db 121 TTAVALPQMIKORSHGVNMSIAGRKVFKGLSVCAKHAHTARSDGLRMEVGKRGIR 180
Qy 186 VGAVLPGPVATLLDDMP-----KAKMEALANGSLMOPLEVAESVLFVTRSKNTVTRDL 241
Db 181 VTICPGAVATIELVHTIDPGTRQMDLATQMTLQGEDIGDTIVFAQAARAHVAVAE 240
Qy 242 VLP 245
Db 241 FVLP 244

RESULT 7
G90050
Hypothetical protein SA2266 [Imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G90050
R:Kutoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G90050
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-231 <KUR>
A:Cross-references: GB:BA000018; PID:g13702428; PIDN:BA843569.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2266
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 26.2%; Score 331; DB 2; Length 231;
Best Local Similarity 31.8%; Pred. No. 3.6e-20;
Matches 75; Conservative 54; Mismatches 99; Indels 8; Gaps 2;

Qy 11 TPLNKGVAATGAAGTGGTCACTLLDAGAKVLLIDREGDKLHKIVAEIGENAVATL 69
Db 2 TVLTDXIAVTVGAGSGIGALATLTHREGAKVVLGRKDKQVANAQADQSVAVVPTD 61
Qy 70 LFNQOVNMLADITELAGLDIFHANAGATGGPVAEGDPDVRVLTNLNNAFRCV 129
Db 62 VTNKEVDELKMAIOATGCGDIIVINSAGOMLSKITTQYQVDEMMDIVNIGTLYTAQ 121
Qy 130 AVLPHMIAORSDDITFTSSAGVAVVIMEPTITASKFAVOAFVITTRQVSGVAVRGA 189
Db 122 AALPMLQSSCHLNIASISGEFVTKSTTYSAKRAVHTITGLEGELATGVKVTSI 181
Qy 190 LPGPVATLLDDMPKAKMEALANGSLMOPLEVAESVLFVTRSKNTVTRDLVILPGSV 245
Db 182 SPGAVDTAITAAYNPSDKK-----LDPDIAEAVLIALTOPKHAVNNEITVAP 230

RESULT 8
T17307

Db 182 VTTPPAINTELLETTIDKETEGMT--SLXKQIGITPDRIASIVAYALDQPEDVNVNEF 239

A; Experimental source

serogroup O1; strain N16961; biotype El Tor

C:Genetics: A:Gene: VC1591 A:Map position: 1 C:Superfamily: rlb1tol dehydrogenase; short-chain alcohol dehydrogenase homology

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 4.96689 Seconds
(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MMNHSVPSMNTPLNGKVAAL.....TRSKNVTVDLILPGSYDL 250

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	87.4	249	1 RIDH_KLEAE	P00335 klebsiella
2	294	23.2	248	1 Y452_LISTIN	Q926K7 listeria in
3	292	23.1	248	1 Y432_LISTMO	P25145 listeria mo
4	284	22.5	246	1 FABG_BACSU	P51831 bacillus su
5	283	22.4	246	1 LINX_THEMA	Q9X248 thermotoga
6	280.5	22.2	250	1 LINX_PSEPA	P50198 pseudomonas
7	273.5	21.6	260	1 YK02_MYCTU	Q10855 mycobacteri
8	271	21.4	238	1 YOXD_BACSU	P14802 bacillus su
9	268.5	21.2	244	1 FABG_VIBCH	Q9K9H7 vibrio chol
10	268.5	21.2	252	1 YDFG_HAEIN	P45300 haemophilus
11	266.5	21.1	244	1 FABG_VIBHA	P55336 vibrio hary
12	264.5	20.9	261	1 DHBA_BACSU	P39071 bacillus su
13	262.5	20.8	248	1 FABG_AOUAE	Q67610 aquifex aeo
14	257	20.3	247	1 FAGI_SYNY3	P73374 synectocyst
15	253	20.0	267	1 HDHA_CLOSO	P50200 clostridium
16	251	19.8	263	1 UCDA_SALTY	P37441 salmonella
17	250.5	19.8	248	1 YDFG_ECOLI	P39831 escherichia
18	250.5	19.8	266	1 PGDH_HUMAN	P15428 homo sapien
19	246.5	19.5	258	1 BDHA_RHIME	O86034 rhizobium m
20	245.5	19.4	261	1 ACT3_STRCO	P16544 streptomyces
21	243.5	19.2	278	1 YALA_RHISM	P55441 rhizobium s
22	243	19.2	258	1 BDHA_ALCEU	Q9X6U2 alcaligenes
23	242.5	19.2	248	1 YDFG_SALTY	P40864 salmonella
24	241.5	19.1	244	1 FABG_SALTY	O85141 salmonella
25	241.5	19.1	320	1 FABG_CUPLA	P28643 cuphea lanc
26	241	19.1	256	1 Y019_THEMA	Q5618 cuphea lanc
27	240	19.0	255	1 YWFD_BACSU	P39640 bacillus su
28	239.5	18.9	247	1 YD50_MYCTU	Q11020 mycobacteri
29	239.5	18.9	263	1 UCDA_ECOLI	P37440 escherichia
30	238.5	18.9	263	1 UCDA_ECOLI	O8X4J4 escherichia
31	237.5	18.8	248	1 FABG_CHLPN	Q928P2 chlamydia p
32	237.5	18.8	262	1 YXBG_BACSU	P46331 bacillus su
33	237	18.7	262	1 VERI_ASPPA	P50161 aspergillus

34	236	18.7	264	1 STCU_EMENT	Q00791 emerichia
35	231.5	18.3	255	1 HDHA_ECOLI	P25529 escherichia
36	231	18.3	267	1 YMT1_YEAST	O05016 saccharomyc
37	230.5	18.2	244	1 FABG_ECOLI	P25716 escherichia
38	230.5	18.2	245	1 NODG_RHIS3	P72332 rhizobium s
39	230.5	18.2	256	1 DH50_RHOSH	Q59787 rhodobacter
40	227.5	18.0	245	1 NODG_RHIME	P06234 rhizobium m
41	226.5	17.9	287	1 HETN_ANASP	P37694 anabaena sp
42	226	17.9	261	1 DHKR_STRCM	P41177 streptomyces
43	225.5	17.8	252	1 YC1K_ECOLI	P31808 escherichia
44	225	17.8	250	1 LINC_PSEPA	P50197 pseudomonas
45	224	17.7	248	1 PHAB_ACISP	P50203 acinetobact

ALIGNMENTS

```

RESULT 1
RIDH_KLEAE          STANDARD:      PRT:      249 AA.
ID  P00335:
AC  21-JUL-1986 (Rel. 01, Created)
DT  01-AUG-1988 (Rel. 08, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Ribitol 2-dehydrogenase (EC 1.1.1.56) (RDH).
GN  RBDT.
OS  klebsiella aerogenes.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX  Klebsiella.
NCBI_TaxID=28451;
RN  (1)
RP  SEQUENCE.
RC  STRAIN=A, and D:
RX  MEDLINE=86050423; PubMed=3904726;
RA  Doehle J.M., Giglio J.R., Moore C.H., Taylor S.S., Hartley B.S.;
RT  "Ribitol dehydrogenase of Klebsiella aerogenes. Sequence and
RT  properties of wild-type and mutant strains.";
RL  Biochem. J. 230:569-578(1985).
RN  (2)
RP  SEQUENCE FROM N.A.
RX  MEDLINE=86050424; PubMed=2933028;
RA  Loviny T., Norton P.M., Hartley B.S.;
RT  "Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the
RT  structural gene.";
RL  Biochem. J. 230:579-585(1985).
RN  (3)
RP  SEQUENCE OF 1-68 FROM N.A.
RX  MEDLINE=85230591; PubMed=3891331;
RA  Wu J., Anderson-Loviny T., Smith C.A., Hartley B.S.;
RT  "Structure of wild-type and mutant repressors and of the control
RT  region of the rib operon of Klebsiella aerogenes.";
RL  EMBO J. 4:1339-1344(1985).
CC  -1- CATALYTIC ACTIVITY: Ribitol + NAD(+) = D-ribulose + NADH.
CC  -1- COFACTOR: BINDS THE COENZYME NICOTINE ADENINE DINUCLEOTIDE.
CC  -1- SUBUNIT: HOMOTETRAMER.
CC  -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN A.
CC  -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC  (SDR) FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL: M25606; AAA25140.1; -
CC  EMBL: X02448; CAA26292.1; -
CC  DR  PIR: A00344; DEKBR.
CC  DR  PIR: S07134; S07134.
CC  DR  PIR: S07135; S07135.
CC  DR  HSSP: P47227; 1BDB.

```

```

DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 20 43 NAD (BY SIMILARITY).
FT ACT_SITE 160 160 BY SIMILARITY.
FT MOD_RES 30 30 CARBOXYMETHYLATION.
FT MOD_RES 126 126 CARBOXYMETHYLATION.
FT VARIANT 196 196 A -> P (IN STRAIN-D).
FT CONFLICT 146 147 AV -> SS (IN REF. 12).
FT CONFLICT 212 212 N -> D (IN REF. 1).
FT CONFLICT 212 212 N -> D (IN REF. 1).
SQ SEQUENCE 249 AA; 26514 MW; C34896C276CC8FEC CRC64;

Query Match
Best Local Similarity 87.4%; Score 1105; DB 1; Length 249;
Matches 215; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

OY 2 MNHVPSPNPLNGKVAATGAAGSIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 61
DB 1 MKHVSVMNNTSLSGKVAATGAAGSIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 60
OY 62 NAVALQDLFENNQVDNMLDIETLAGGLDIFHANAGAYIGGPAEGDPVMDVRLNINI 121
DB 61 NAFALQVDLMDQADVNDLQGLIQLTGRDIFHANAGAYIGGPAEGDPVMDVRLNINI 120
OY 122 NAARCRVAIVPHMIAQRSGDIIFTSSAGVVPYIMEPIYASKFVAQAFVHTTRROYSQ 181
DB 121 NAAARCRVSVPLHLIAQRSGDIIFTAVIAGVVPYIMEPIYASKFVAQAFVHTTRROYAQ 180
OY 182 YGVAVGAVLPGPVYATALLDMPKAKMEBALANGSLMOPIEVASVLEFVTRSKVNTYVDL 241
DB 181 YGVAVGAVLPGPVYATALLDMPKAKMEBALANGSLMOPIEVASVLEFVTRSKVNTYVDI 240
OY 242 VILPGVDL 250
DB 241 VILPNSVDL 249

RESULT 2
Y452_LISIN STANDARD; PRT; 248 AA.
ID Y452_LISIN
AC Q92EK7;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical oxidoreductase Lin0452 (EC 1.-.-.-).
GN LMO0452.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Eutlier K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapatk G.,
RA Madueno E., Maltounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species.";
RA Science 294:849-852(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL596165; CAC95684.1; -.
DR ListList: LMO0452; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26808 MW; EAEB01F110306F CRC64;

Query Match
Best Local Similarity 23.2%; Score 294; DB 1; Length 248;
Matches 77; Conservative 31.6%; Pred. No. 4.1e-17;
Matches 77; Conservative 48; Mismatches 105; Indels 14; Gaps 5;

OY 13 LNKGVAAITGAASIGLQCAKTLIDAGAKVLLIDREGDKLHKIV---AEIGENAVLQL 68
DB 3 LKNVVIITGASGIGETATALLAEKGAKLVAARVEKLEKIVQTIKASSGEAIFA-KT 61
OY 69 DLFENNQVDNMLDIETLAGGLDIFHANAGAYIGGPAEGDPVMDVRLNINIARCV 128
DB 62 DVTRREDNKKLVLEATIEYGVADAIPLNAGIMPSPLALKEDWEDQWIDINIKVNLGI 121
OY 129 RAVLPHMIAQRSGDIIFTSSAGVVPYIMEPIYASKFVAQAFVHTTRROYSG--VRV 186
DB 122 AAVLPFSLAQSGHIIATSSVAGLKAIVGAVYATKMAVNDLMEVLRMEASQGTNIRT 181
OY 187 GAVLPGPVYATALLDMPKAKMEBALAN----GSLMOPIEVASVLEFVTRSKVNTYVDL 241
DB 182 ATIVPAIINTELLETITDKETEGQMTNLKKQYG--VTPDRIASIVAAVQDPEDINVEF 239
OY 242 VILP 245
DB 240 TVGP 243

RESULT 3
Y432_LISMO STANDARD; PRT; 248 AA.
ID Y432_LISMO
AC P25145;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical oxidoreductase Lmo0432 (EC 1.-.-.-) (ORFA).
GN LMO0432.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=1905979;
RA Gaillard J.-L., Berche P., Frehel C., Gouin E., Cossart P.,
RT "Entry of L. monocytogenes into cells is mediated by internalin, a
RT coccid protein remnant of surface antigens from Gram-positive
RT cocci.";
RT Cell 65:1127-1141(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Eutlier K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapatk G.,
RA Madueno E., Maltounam A., Mata Vicente J., Ng E., Nedjari H.,

```



```

RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlunetter T., Simoes N., Tierrez A.,
RA Vazquez-Poland J.-A., Voss H., Wehlund J., Cossart P.,
RL Science 294:849-852(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M67471, AAA25288.1, -.
DR EMBL, AL591975, CAC98511.1, -.
DR PIR, B39930, B39930.
DR HSSP, O70351, 1E6W.
DR ListList: LMO000432, -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short.1.
DR PROSITE, PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 9 33 NAD OR NADP (By SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
FT CONFLICT 48 48 T -> I (IN REF. 1).
FT CONFLICT 182 182 V -> A (IN REF. 1).
SQ SEQUENCE 248 AA; 26821 MW; 50262487D29FD935 CAC64;

Query Match 23.1%; Score 292; DB 1; Length 248;
Best Local Similarity 32.1%; Pred. No. 6e-17;
Matches 79; Conservative % 47; Mismatches 106; Indels 14; Gaps 5;

QY 13 LKNGVAATGAAGSGGLGCARTLLDAGAKVYLIDEGSKLHKY----AEIGENMYAQL 68
DB 3 LKKNVITITGSSSGIGKATALLAKGKVLAAARVERLEIVOTIKANSGEALFA-KT 61
QY 69 DLFNNQOVNMLADIIELAGLDIFHANAGAYTGGPVAEGDPDWDVRLNLINMAFRCV 128
DB 62 DVTREKEDKKLVELAIEERYGVDAIFLNAIGIMPNSPLSKLKEDEWEQMDINIKVLNGI 121
QY 129 RAVLPHMLAQSGLDIFFTSSIAGVVPVIMEPTLYTASKFAVCAFEVTTTRQVSQYG--VAV 186
DB 122 AAVLPSEFIAQSSGHIATSSVAAGLKAYPGAGVYATKNAVRDLMVLRRESAQECTNTRT 181
QY 187 GAVLPGPVVTALLDMPAPKAKEALANGSL-----MPLVEASVLEFWTRSKNTVRDL 241
DB 182 VTIIPAALNTELELTITTKTEQGMFT--SLVKQYQITTDRIASIVAYALDQEDVNVNNEF 239
QY 242 VILPES 247
DB 240 TVGPFS 245

RESULT 4
FABG_BACSU STANDARD: PRT; 246 AA.
AC P51831: 031733;
AC P51831: 1996 (Rel. 34, Created)
DT 15-JUN-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl)-
DE acyl carrier protein reductase).
DE FABG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RX [1]
RX SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=96326321; PubMed=8759840;
RX

```

RA Morbidon H.R., de Mendoza D., Cronan J.E. Jr.;
RT "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RL lipid biosynthesis genes";
RN J. Bacteriol. 178:4794-4800(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertlan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallitz E.J., Grand G.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golligly E., J. Grand G.,
RA Giuseppe G., Guy B.J., Haga K., Hatch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manel C., Medigue C.,
RA Medina N., Meliade R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Moone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Poll T.M., Portetelle D., Portoullik S., Prescott A.M.,
RA Priesecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche A., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schreier R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambolt R., Wedler E., Weller H., Weltzieneger T.,
RA Winters P., Wipit A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RN Nature 390:249-256(1997).
[3]
RP SEQUENCE OF 1-172 FROM N.A.
RC STRAIN-168:
RX MEDLINE-98195738; PubMed-9534248;
RA Foulger D., Errington J.;
RT "A 28 kbp segment from the spoVM region of the Bacillus subtilis 168
RT genome";
RN Microbiology 144:801-805(1998).
[4]
RP SEQUENCE OF 230-246 FROM N.A.
RC STRAIN-168:
RA Oguero A., Kakeshita H., Takamatsu H., Nakamura K., Yamane K.;
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: [3R]-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U59433; AAC44307.1; -
DR EMBL: Z99112; CAB13464.1; -
DR EMBL: Y13937; CA874250.1; -
DR EMBL: D64116; BAI0974.1; -
DR HSSP: Q12634.1YBV.
DR Subtilist; Bc11535; fabG.

```

DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; FALSE_NEG.
KM Fatty acid biosynthesis; Oxidoreductase; NADP: Complete proteome.
FT ACT_SITE 9 33 NADP (BY SIMILARITY).
FT CONFLICT 154 154 D -> A (IN REF. 1).
FT SEQUENCE 246 AA; 26282 MW; C6A391167D3237DC CRC64;
SQ
Query Match 22.5%; Score 284; DB 1; Length 246;
Best Local Similarity 31.8%; Pred. No. 2.7e-16;
Matches 78; Conservative 47; Mismatches 110; Indels 10; Gaps 5;
QY 13 LKGVAAITGAASGIGLQCAKTLIDAGAKVVLIDREGD--KLHKIVAE--LGENAYALQ 67
DB 2 LNDKTAIVTGAASRGISALDLAKSGANVY-VNYSGEAKANEVDEIKSMGRKALAVK 60
QY 68 LDLENNQVDMMLADIIELAGGLDIFHANAGAYIGGPAEGDPVWDRLVNLINAAFC 127
DB 61 ADVNPEDVNMKIKETLSVFSTIDILVNMAGITRDNLIRKKEDEMDVIVINLKGVFNC 120
QY 128 VRAVLPHMIAGRSDDIIFTSIAGVVPVIMEPIYTASKEFAVQAFVHTTRQVSGYRVG 187
DB 121 TKAVTRQMKKORSRIITVSSIVGSGNPGQANTVAAGVIGLTKSSAKELASRNITVN 180
QY 188 AVLPGPVYVALLDDMPKAKMEALANGSLM--OPIEVAESVLEPMVTR-SKNVTYRDLVI 243
DB 181 AIAAGFIETDPTDKLADVQDEMLKQIPLARFGEPSDVSVVTFLASGAYVMGQTLHI 240
QY 244 LPGSV 248
DB 241 DGMV 245
RESULT 5
FABG_THEME STANDARD: PRT; 246 AA.
AC Q9248;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-
acyl carrier protein reductase).
GN FABG OR TM1724.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; Pubmed=10360571;
RA NELSON K.E., Clayton R.A., GILL S.R., GWIN M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utlacker T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Selzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 359:323-329(1999).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001811; AAD36790.1; -.
DR HSSP: P50162; IAE1.
DR TIGR: TM1724; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KM Fatty acid biosynthesis; Oxidoreductase; NADP: Complete proteome.
FT ACT_SITE 10 34 NADP (BY SIMILARITY).
FT NP_BIND 154 154 BY SIMILARITY.
FT SEQUENCE 246 AA; 26401 MW; BC08904D28099142 CRC64;
SQ
Query Match 22.4%; Score 283; DB 1; Length 246;
Best Local Similarity 30.7%; Pred. No. 3.2e-16;
Matches 75; Conservative 47; Mismatches 112; Indels 10; Gaps 4;
QY 13 LKGVAAITGAASGIGLQCAKTLIDA--GAKVVLIDREGDKLHKIVAE--LGENAYALQ 67
DB 3 LKGVCLITGAASIGK--ATTLFPAQEGATVINGDISKEVDSLVAEGLPKVDPVY 60
QY 68 LDLENNQVDMMLADIIELAGGLDIFHANAGAYIGGPAEGDPVWDRLVNLINAAFC 127
DB 61 LVNVDROQIEKVEKVVQKGRIDVLVNMAGITRDALLVRKKEDEMDVAVINLKGVFNC 120
QY 128 VRAVLPHMIAGRSDDIIFTSIAGVVPVIMEPIYTASKEFAVQAFVHTTRQVSGYRVG 187
DB 121 TVMVPYPIKORNGSIVNVSSVGIYGNPGQNTVAASKAGVIGTKTWAELASRNIRVN 180
QY 188 AVLPGPVYVALLDDMPKAKMEALAN--GSLMOPIEVAESVLEPMVTRSKNVTYRDLVI 244
DB 181 AVAGFIETPTETKPLKAEFTALSRIPLGRFGEPAVQVILFLADESSVYTGQYIGI 240
QY 245 PGSV 248
DB 241 DGCL 244
RESULT 6
LINK_PSEPA STANDARD: PRT; 250 AA.
AC P50198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-)
DE (2,5-DDOL dehydrogenase).
GN LINK.
OS Pseudomonas paucimobillis (Sphingomonas paucimobillis).
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=13689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UT26;
RX MEDLINE=94252977; Pubmed=7515041;
RA Nagata Y., Ohtomo R., Miyauchi K., Fukuda M., Yano K., Takagi M.;
RT "Cloning and sequencing of a 2,5-dichloro-2,5-cyclohexadiene-1,4-diol
RT dehydrogenase gene involved in the degradation of gamma-
RT hexachlorocyclohexane in Pseudomonas paucimobillis.";
RL J. Bacteriol. 176:3117-3125(1994).
CC -1- FUNCTION: DEGRADATION OF 2,5-DICHLORO-2,5-CYCLOHEXADIENE-1,4-DIOL
(2,5-DDOL) INTO 2,5-DICHLOROHYDROQUINONE (2,5-DCHQ). LINK IS NOT
ESSENTIAL TO GAMMA-HCH DEGRADATION.
CC -1- PATHWAY: DEGRADATION OF GAMMA-HEXACHLOROCYCLOHEXANE.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D23722; BAA04939.1; -
 DR HSSP: P19992; 1HDC.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT.1.
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
 FT NP_BIND 9 NAD (BY SIMILARITY).
 FT ACT_SITE 156 156 BY SIMILARITY.
 SQ SEQUENCE 250 AA; 25492 MW; 8C52703FF76382CF CRC64;

Query Match 22.2%; Score 280.5; DB 1; Length 250;
 Best Local Similarity 31.0%; Pred. No. 5.3e-16;
 Matches 75; Conservative 47; Mismatches 113; Indels 7; Gaps 3;

OY 9 MNTPLNGKVAATGASGIGLCAKTLIDGAKVVLIDREGDKLHKIVAE--GENAYA 65
 DB 1 MANRLAGKVALITGASGLGAQAQKRAEGAKVAVIGLNEEAKGVAEIRAGGDALF 60
 OY 66 LQDLFFNQVDNMLADIETELAGLDIFHANAGAYIGPVAEGDPVMDRVNLINAAE 125
 DB 61 ILVDYDASWNNATIAAVADGEGGLTTLSTAGIIHPGFEESIGMKMNAVNTAIF 120
 OY 126 KCVRAVLPMLAORSGLDIFTSSIAGVVPVIMPEITASKFAVAQAVHTTRQVSGYR 185
 DB 121 LGIKAIPELVKSGNSIINISLIGMPTAGNASCATKAARIMSKAAALEFVGRGVA 180
 OY 186 VGVAVLPGPVVTLDDMPKAKMEALAN---GSLMQPIEVAESVLPMT-RSKNVTVRDL 241
 DB 181 VMTIVPGMNTPTTANVPDVLKQRTSQIIPMGKLPIDIANGLPLASDEAKYITGVDL 240
 OY 242 VI 243
 DB 241 PI 242

RESULT 7
 YK02_MYCTU STANDARD; PRT; 260 AA.
 ID YK02_MYCTU
 AC Q10855;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative oxidoreductase RV2002 (EC 1.-.-.-).
 GN FABG3 OR RV2002 OR MT2058 OR MTCY39.16C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmler K., Parkhill J., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Horsby T., Jagels K., Krogh A., McLaren J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: Z74025; CAA98414.1; -
 DR EMBL: AE007057; AAK46335.1; -
 DR HSSP: P19992; 1HDC.
 DR TIGR: MT2058; -
 DR Tuberculist: RV2002; -
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT.1.
 KW Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
 FT NP_BIND 11 35 NAD (BY SIMILARITY).
 FT ACT_SITE 153 153 BY SIMILARITY.
 FT CONFLICT 174 174 S -> G (IN REF. 2).
 SQ SEQUENCE 260 AA; 27030 MW; 0935A1ED36220B7 CRC64;

Query Match 21.6%; Score 273.5; DB 1; Length 260;
 Best Local Similarity 31.6%; Pred. No. 2.1e-15;
 Matches 77; Conservative 47; Mismatches 111; Indels 9; Gaps 4;

OY 9 MNTPLNGKVAATGASGIGLCAKTLIDGAKVVLIDREGDKLHKIVAELEGNAAYA 65
 DB 1 MSGRLLGKVALVSGARGKASHVRAMVAGAVVFCGLIDDEGS---KAAALADAAAY 57
 OY 66 LQDLFFNQVDNMLADIETELAGLDIFHANAGAYIGPVAEGDPVMDRVNLINAAE 125
 DB 58 VHLDTQPAQWMTAAVDTAFAFGGLHVLVNNAGILNIGTIEDYALTEWQRIIDVNLTV 117
 OY 126 KCVRAVLPMLAORSGLDIFTSSIAGVVPVIMPEITASKFAVAQAVHTTRQVSGYR 185
 DB 118 LGIRAVVKKPKKRGSGSIINISIEIGACTVACHGYATKFAVAGLTKSTALEGPSGR 177
 OY 186 VGVAVLPGPVVTLDDMPKAKMEALANGSLMQPIEVAESVLPMT-RSKNVTVRDLVIL 244
 DB 178 VNSIHGGLYKTPMTDWPEDIFOTAL--GRAAPVEVSNLYVLADDESSYSGAFYVD 235
 OY 245 PGSV 248
 DB 236 GGTV 239

RESULT 8
 YOXD_BACSU STANDARD; PRT; 238 AA.
 ID YOXD_BACSU
 AC P14802;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical oxidoreductase yoxd (EC 1.-.-.-).
 GN YOXD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN-168;
RX MEDLINE-91192601; PubMed-1849493;
RA Ann K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W33 chromosomes.";
RL Gene 98:107-112(1991).
RN [2]
RP SEQUENCE OF 62-238 FROM N.A.
RX STRAIN-168;
RX MEDLINE-88040469; PubMed-1118336;
RA Carrigan C.M., Haarsma J.A., Smith M.T., Wake R.G.;
RT "Sequence features of the replication terminus of the Bacillus
RT subtilis chromosome.";
RL Nucleic Acids Res. 15:8501-8509(1987).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06168; CAA29533.1; -
CC EMBL: Z99114; CAB13743.1; -
CC PIR: S01270; S01270.
CC HSP: O70351; 1E3S.
CC Subtilist: Bg11048; YOXD.
CC Interpro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Hypothetical protein; Oxidoreductase: Complete proteome.
CC NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 155 155 BY SIMILARITY.
CC SEQUENCE 238 AA; 25299 MW; 9CCTA8D1204DF248 CRC64;
SQ
Query Match 21.4%; Score 271; DB 1; Length 238;
Best Local Similarity 32.5%; Pred. No. 3e-15;
Matches 74; Conservative 37; Mismatches 101; Indels 16; Gaps 4;
QY 13 LNCVAAITGAASIGIGQCAKTLIDAGAKVYLIDREGDKLHKIYAE--IGENAYALQLD 69
DB 4 LQHTTALTITGGRGIGATLALAKKEGVNIGLIGRTSANEKVAEEKVALVKAFAAD 63
QY 70 LFNNOGVNMLADITIELAGLDIFHANAGAYIGGPAVEGDDVDVRLNINNAFCVR 129
DB 64 VKDADQVNOVAOVKEQLGIDIDILINNAIGSKFEGFLDSADEMENTIYVNLKGVHYTR 123
QY 130 AVLPHTIAQRSGDITFTSSIAGVVPVIMETPTASKRAVOAFVHTTRROYSOYGVRCAY 189
DB 124 AVLEPMIERKAGDIINISFVAGGCAVTSYASKFAVIGLSTELMQEVRKHINIRVAL 183
QY 190 LFGVYVATLLDDWPKAKMEALANGS---LMQPIEVAESVLEFWYTRSK 234
DB 184 TPSTVASDM-----SIEINLTLDGNEPKVWQPEDLAE---YVVAOLK 221

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-El Tor N16961 / Serotype O1;
RX MEDLINE-20406833; PubMed-10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tellein H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -----
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF004276; AAF95169.1; ALT_INT.
CC HSP: P18992; 1HDC.
CC TIGR: VC2021; -.
CC Interpro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Fatty acid biosynthesis; Oxidoreductase: NADP. Complete proteome.
CC NP_BIND 10 34 NADP (BY SIMILARITY).
CC ACT_SITE 151 151 BY SIMILARITY.
CC SEQUENCE 244 AA; 25566 MW; 9FB2E8278D7CC3CE CRC64;
SQ
Query Match 21.2%; Score 268.5; DB 1; Length 244;
Best Local Similarity 32.3%; Pred. No. 4.9e-15;
Matches 72; Conservative 39; Mismatches 107; Indels 5; Gaps 2;
QY 13 LNCVAAITGAASIGIGQCAKTLIDAGAKVYLIDREGDKLHKIYAEIGENAYALQLD 72
DB 3 LEGKVALVTGASRGIGAKIAIELLAERGAIVIGTSSGAAISDYLDGNGKGMALVTN 62
QY 73 NQOVNMLADITIELAGLDIFHANAGAYIGGPAVEGDDVDVRLNINNAFCVRAVL 132
DB 63 PESIEAVLKAITDEFQVDILVNNAGITRDNLARMKEEESDIMEETNLTISIFRLSKAVL 122
QY 133 PHMTAQRSGDITFTSSIAGVVPVIMETPTASKRAVOAFVHTTRROYSOYGVRCAYLPG 192
DB 123 KGMKKRKGRIITNVGSVVGITGNAGQANFYAAKAGVIGFTKSMAREVASRGVTVNTYAPG 182
QY 193 PVVT-----ALDDWPKAKMEALANGSLMOPIEVAESVLEFWYTRSK 231
DB 183 FIETDMTKALNDEQRTATLQVPA-GRLDGDPREIATASAVAFLAS 224

```

RESULT 9
FABG_VTBCH
ID FABG_VTBCH STANDARD; PRT; 244 AA.
AC 09KOH7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase).
FABG OR VC2021.
GN
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;

RESULT 10
YDFG_HAEIN
ID YDFG_HAEIN STANDARD; PRT; 252 AA.
AC PA5200;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical oxidoreductase H11430 (EC 1.-.-.-).
GN H11430.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

```

OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Genem C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. STRONG, TO OTHER BACTERIAL HOMOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32822; AAC23077.1; -.
CC DR HSSP: P50162; IAE1.
CC DR TIGR: H11430; -.
CC DR InterPro: IPR002198; ADH_short.
CC DR Pfam: PF00106; adh_short.1.
CC DR PRINTS: PR00080; SDRFAMILY.
CC DR PROSITE: PS00061; ADH_SHORT, 1.
CC NP_BIND 7 31 NAD OR NADP (BY SIMILARITY).
CC FT ACT_SITE 150 150 BY SIMILARITY.
CC SQ SEQUENCE 252 AA; 27398 MW; 227FB28CE7D9A98 CRC64;

Query Match 21.2%; Score 268.5; DB 1; Length 252;
Best Local Similarity 31.8%; Pred. No. 5, 1e-15;
Matches 75; Conservative 41; Mismatches 109; Indels 11; Gaps 3;

OY 18 AAITGAAGTIGLQCAKTLIDAGAKVYLIDREGDKLHKYAELEGNAVALQDLDFNNQVD 77
DB 6 ALVTGATAGFLAICKKLEAGKVIKGTGRADRLAEIHSQGNFLPLAFDIRDOAT 65
OY 78 NMLADIIELAGLIDIFHANAGAVIG-GPAEGDPDWDVRLNLINAAEFRCVRAVLP 136
DB 66 NALNTLPEQWQAVDLVNNAGLALGLEPAHKADLDQWYOMIDINITGLVITRLVLP 125
OY 137 AQRSGDIIFTSSIAGVPIWEPIYASKFAVAQAFVHTTRROYSOYGVAVLP 192
DB 126 ARNYGQIIMLSIAGTYPAAGSNVYGTAKFVTOFSLNLRADLAGKIRVSNVEPGCG 185
OY 193 ---PVYFALLDMPKAKMEALANGSLMOPIVEAESVLEPMVTRSKVNTYRDVILP 245
DB 186 TEFNSVNRHFHGDDEAKAYE---NWSQVOPEDIANIVLWLEHOOPEHVNIIRLEVP 238

RESULT 11
FABG_VIBHA STANDARD: PRT; 244 AA.
AC P55336;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-
DE acyl carrier protein reductase).
DE FABG.
OS Vibrio harveyi.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Byers D.M.;
RT "Isolation of Vibrio harveyi acyl carrier protein and the fabg, acpP,
RT and fabg genes involved in fatty acid biosynthesis.";
RL J. Bacteriol. 178:571-573(1996).
CC -1 CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NAD(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1 PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39441; AAC43589.1; -.
CC DR HSSP: P19992; IHDC.
CC DR InterPro: IPR002198; ADH_short.
CC DR Pfam: PF00106; adh_short.1.
CC DR PRINTS: PR00080; SDRFAMILY.
CC DR PROSITE: PS00061; ADH_SHORT, 1.
CC KW Fatty acid biosynthesis; Oxidoreductase; NADP.
CC FT NP_BIND 10 34 NADP (BY SIMILARITY).
CC FT ACT_SITE 151 151 BY SIMILARITY.
CC SQ SEQUENCE 244 AA; 25519 MW; FC41A1C65B8CDA9A CRC64;

Query Match 21.1%; Score 266.5; DB 1; Length 244;
Best Local Similarity 32.3%; Pred. No. 7, 2e-15;
Matches 72; Conservative 39; Mismatches 107; Indels 5; Gaps 2;

OY 13 LNSKVAITGAAGTIGLQCAKTLIDAGAKVYLIDREGDKLHKYAELEGNAVALQDLDFN 72
DB 3 LKGIKALVTGASGIGIRALIAELLVERGATVIGTATSGCAAAISEYLEGNGKGLALNVT 62
OY 73 NQOVDNMLADIIELAGLIDIFHANAGAVIGGPAEGDPDWDVRLNLINAAEFRCVRAV 132
DB 63 VESIEATLKTINDECAIDIDLNNAGITRDNLIMRKDDDEMDINDINLTPITRMSKAVL 122
OY 133 PHMIAQRSGDIIFTSSIAGVPIWEPIYASKFAVAQAFVHTTRROYSOYGVAVLP 192
DB 123 RGMKKRRAGRIINVSIVGTMGNAQGTNYAAAGVIGFTKSMAREVAGSVYNTVAPG 182
OY 193 PVYF---ALLDMPKAKMEALANGSLMOPIVEAESVLEPMV 231
DB 183 FIETDMTKALINDQRAATLSNVPA-GRLGDPREIASAVVFLAS 224

RESULT 12
DHBA_BACSU STANDARD: PRT; 261 AA.
AC P39071;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28)
DE (Cold shock protein CS14).
DE DHBA OR ENTA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;

```

```

RX MEDLINE-97080557; PubMed-8921902;
RA Rowland B.M., Grossman T.H., Osburne M.S., Taber H.W.;
RT "Sequence and genetic organization of a Bacillus subtilis operon
RL encoding 2,3-dihydroxybenzoate biosynthetic enzymes.";
RN Gene 178:119-123(1996).
RP SEQUENCE OF 59-255 FROM N.A.
RC STRAIN-168;
RX MEDLINE-94040785; PubMed-8224884;
RA Adams R., Schumann W.;
RT "Cloning and mapping of the Bacillus subtilis locus homologous to
RL Escherichia coli ent genes.";
RN Gene 133:119-121(1993).
RP SEQUENCE OF 1-11.
RC STRAIN-168 / JH642;
RA Graumann P.L., Schmid R., Marahel M.A.;
RL Submitted (OCT-1997) to the SWISS-PROT data bank.
CC -1- CATALYTIC ACTIVITY: 2,3-dihydro-2,3-dihydroxybenzoate + NAD(+) =
CC 2,3-dihydroxybenzoate + NADH.
CC -1- PATHWAY: 2,3-dihydroxybenzoate biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: IN RESPONSE TO LOW TEMPERATURE.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U26444; AAC4630.1; -
DR EMBL: L08644; AA16899.2; -
DR EMBL: 299120; CAB5190.1; -
DR PIR: P06883; P06883.
DR HSSP: P08074; 1CYD.
DR Subtilisin: BG11019; dba.
DR Interpro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD; Complete proteome.
FT NP_BIND 12 36 NAD (BY SIMILARITY).
FT ACT_SITE 157 157 P -> D (IN REF. 2).
FT CONFLICT 146 146 IDA -> MRC (IN REF. 2).
FT CONFLICT 231 234 IADA -> MRC (IN REF. 2).
FT CONFLICT 247 256 TMHNLGVGG -> RCIFMRCAT (IN REF. 2).
SQ SEQUENCE 261 AA; 27494 MW; 00B0E8BA53AB407C CRC64;

Query Match 20.9%; Score 264.5; DB 1; Length 261;
Best Local Similarity 28.6%; Pred. No. 1,1e-14;
Matches 72; Conservative 48; Mismatches 115; Indels 17; Gaps 4;

QY 13 LNCNVAITGAASGIGLOCAKTLIDAGAKVLLIDREDKILHIAEL---GENAYALQD 69
DB 6 IEGIAITITGAAGSIGCAVAVARTLASGCAHIAVDYNEPEKLEKVSLSKAERHAEAPAD 65
QY 70 LFNNOQVDMNLADIIELAGLIDIFHANAGAYIGGPAEGDPVDVRYLNLINAAFCVR 129
DB 66 VRDAAIDEITARIEREMGPIDILVNVAAGVLRPGLHLSLSEDEAEVFSVSTGVFNASR 125
QY 130 AVLPHMAQSSGDIIFTSIAGVVPVIMEPIYTSKFAVOAFVHTTRQVSQIGVRGAV 189
DB 126 SVSKYMDRRSGSIYTVGSNPAGVPRTSMAYASSKAAAVMTKCLGLELAETVIRCNIV 185
QY 190 LQGVYVAL-LDMPKAKMEALANGS-----LMQIEVAESYLEFMT-RSKN 235
DB 186 SPGSTEDMOMSLMADENGAEOYIKGSLTFKTCIGIPKLKLAIPSDIADAVLELVSGAGH 245
QY 236 VIVRDLVILPGS 247
DB 236 VIVRDLVILPGS 247

```

```

DB 246 ITMHNLCVDGGA 257
RESULT 13
FABG_AOUAE
ID FABG_AOUAE STANDARD; PRT; 248 AA.
AC 067610;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (Ec 1.1.1.100) (3-ketoacyl-
DE acyl carrier protein reductase).
GN FABG OR AO.1716.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE-98196666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000752; AAC07575.1; -
DR HSSP: P50163; 2AE1.
DR Interpro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT NP_BIND 12 36 NADP (BY SIMILARITY).
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26867 MW; 5CFD9EB9ADB3F2C5 CRC64;

Query Match 20.8%; Score 262.5; DB 1; Length 248;
Best Local Similarity 26.6%; Pred. No. 1,6e-14;
Matches 67; Conservative 56; Mismatches 112; Indels 17; Gaps 4;

QY 9 MNTPLNKVAITGAASGIGLOCAKTLIDAGAKVLLIDREDKILHIAEL---GENAY 64
DB 1 MEIKLGKVSIVYTGSGIGRAIAEKLASAGSTVITIGTSGERAKAAVEIANKYGVAKH 60
QY 65 ALQDLDFNNOQVDMNLADIIELAGLIDIFHANAG-----AYIGGPAEGDPVDVRYLNL 119
DB 61 GVENMLLSEESINKAFEEITNLDGIDILVNNAGITDOKLEFLRNSLSD-----WEEVLKV 115
QY 120 NINAFCVRAVLPHMAQSSGDIIFTSIAGVVPVIMEPIYTSKFAVOAFVHTTRQVSQ 179
DB 116 NLITGTFVLTVNSLAKMKQKRGRIYVNSSVYGFNGGVNNYSTTKAGLIGFTSLAKEL 175
QY 180 SQYGVRRGAVLPGFV---VTALLDDMPKAKMEALANGSLMQIEVAESYLEFMT-RSKN 236
DB 176 APRNVLVNAVAPGFIETDMNAVISEEIKQYKKEQIPICGRGSPREAVANVYLFCSELASY 235

```

Oy 237 TVRDLVLPQSV 248
 Db 236 ITGEVTHVNGM 247

RESULT 14

FAGL_SYNY3 STANDARD: PRT; 247 AA.
 ID FAGL_SYNY3
 AC P73574;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 3-oxoacyl-[acyl-carrier protein] reductase 1 (EC 1.1.1.100) (3-oxoacyl-acyl carrier protein reductase 1).
 GN FAGL OR SLR0886.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., MiyaJima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RA "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1 CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC -1 PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: D90907; BAA17614.1; -
 DR HSP: P50162; IAE1.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 11 35 NADP (BY SIMILARITY).
 FT ACT_SITE 156 156 BY SIMILARITY.
 FT SEQUENCE 247 AA; 25724 MW; 91BEP9409C777FE20 CRC64;

Query Match 20.3%; Score 257; DB 1; Length 247;
 Best Local Similarity 30.2%; Pred. No. 4, 4e-14;
 Matches 68; Conservative 46; Mismatches 107; Indels 4; Gaps 2;

Oy 11 TPUNGVAITGAASGIGQCAKTLIDAGAKVYL-IDREGDKLHKIVAEI--GENAVAL 66
 Db 2 TALTNOVALVTGASRGKATLALATGMKVVYNAOSTADAVVAELIANGGALAV 61
 Oy 67 QLDLENNQOVDMADIEFLAGGLDIFHANAGAYIGGPAESDPPVDVRLNININAAR 126
 Db 62 QANVANADEVDOLIKITLTKFSRIDLVNNAAGITTRDTLLRMKLEDMQAVIIDINTLGEVL 121
 Oy 127 CVRAVLPHTIAQRSGDIIETSSIAGVVPVWEIYATSKFAVOAFHTTRROVSQGVAV 186
 Db 122 CTAAVSKMLKGRSGRIINTSVAGMMGNPDGAANNSSAAGVYGFRTYAKKELASRGVTV 181
 Oy 187 GAVLPGPVVTALDDMPKAKMEBALNGSLMOPIEVAESVLPVTV 231

Db 182 NAVAPGFIATDMTENLNAEPILOFIPLARYGOPEEVAAGTIRFLAT 226

RESULT 15

HDHA_CLOSO STANDARD: PRT; 267 AA.
 ID HDHA_CLOSO
 AC P50200;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 7-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) (bile acid 7-dehydroxylase) (7-alpha-HSDH).
 OS Clostridium sordellii.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OX Clostridium.
 OX NCBI_TaxID=1505;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
 RX STRAIN=ATCC 9714;
 RX MEDLINE=94327451; PubMed=8050999;
 RA Coleman J.P., Hudson L.L., Adams M.J.;
 RA "Characterization and regulation of the NADP-linked 7 alpha-hydroxysteroid dehydrogenase gene from Clostridium sordellii.";
 RL J. Bacteriol. 176:4865-4874(1994).
 CC -1 CATALYTIC ACTIVITY: 3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholanoate + NADP(+) = 3-alpha,12-alpha-dihydroxy-7-oxo-5-beta-cholanoate + NADPH.
 CC -1 SUBUNIT: HOMOTETRAMER (PROBABLE).
 CC -1 INDUCTION: BY BILE ACID. EXPRESSED DURING GROWTH-PHASE.
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDS) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: I12058; AAA53556.1; -
 DR HSP: P25529; IAH1.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR KW Oxidoreductase; NADP; Bile acid catabolism.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 158 158 BY SIMILARITY.
 FT SEQUENCE 267 AA; 29177 MW; 8560207B9EDCE0E CRC64;

Query Match 20.0%; Score 257; DB 1; Length 267;
 Best Local Similarity 31.2%; Pred. No. 1e-13;
 Matches 72; Conservative 37; Mismatches 106; Indels 16; Gaps 5;

Oy 13 LNKVAITGAASGIGQCAKTLIDAGAKVYLIDREGDKLHKIVAEI--GENAVAL 72
 Db 4 LENKVALVTSATRGIGLASIKLAQNGAIVYKWRRLAEQTQCDYKKEGLTKLPFFD 63
 Oy 73 NCOVD---NMLADIIELAGGLDIFHANAGAYIGP-----VAGDDPVDVRLNININA 124
 Db 64 AYVIDITKEMIDITJINKEGKIDLVNNGF--GRPEKDLVNGDEDTPEELFNVNGSV 121
 Oy 125 FRCVRAVLPHTIAQRSGDIIETSSIAGVVPVWEIYATSKFAVOAFHTTRROVSQGVAV 184
 Db 122 YRLSKLIIPIHMEINKGSIYINISVGSIPDISRIGVGSKSVNNITKOIAIQVAKYGI 181
 Oy 185 RQAVVLPGPVVT-----ALLDDMPKAKMEBALNGSLMOPIEVAESVLPVTV 230
 Db 182 RCNAVLPGLTADDAANSMPEDEFRKSFVLSHPLN-RIGNEDIANSVLPFV 231

Tue Mar 25 12:01:56 2003

us-09-802-208b-3.rsp

Page 10

Search completed: March 13, 2003, 16:53:42
Job Time : 6.96689 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:43:37 ; Search time 15.9354 Seconds
(without alignments)
3232:351 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265
Sequence: 1 MMHNSVPSMNTPLNGKVAAL.....TRSKNVTYRDVILFSGVDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*
15: sp-virus:*
16: sp_bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	2 Q9F4L7	Q9F4L7 escherichia
2	843	66.6	242	16 Q8YBC0	Q8YBC0 bruceella me
3	573	45.3	243	16 Q8U7X6	Q8U7X6 agrobacteri
4	366	28.9	247	2 Q9LCV7	Q9LCV7 streptomyce
5	342	27.0	247	2 Q54197	Q54197 streptomyce
6	333	26.3	249	16 Q930M2	Q930M2 rhizobium m
7	331	26.2	231	16 Q99RPS	Q99RPS staphylococ
8	317	25.1	248	2 Q9RH22	Q9RH22 zymomonas m
9	309	24.4	325	4 Q9BTF9	Q9BTF9 homo sapien
10	304	24.0	325	4 Q9UFM6	Q9UFM6 homo sapien
11	303.5	24.0	253	16 Q92NFB	Q92NFB rhizobium m
12	298	23.6	248	2 Q9LA07	Q9LA07 paenibacill
13	295.5	23.4	246	2 Q9EX74	Q9EX74 rhodococcus
14	295.5	23.4	255	1 Q34187	Q34187 halobacteri
15	292	23.1	248	2 Q9EXG3	Q9EXG3 listeria mo
16	291.5	23.0	258	16 Q982J8	Q982J8 rhizobium 1

17	289	22.8	247	16 Q92AK1	Q92AK1 listeria in
18	288.5	22.8	252	16 Q9KRP5	Q9KRP5 vibrio chol
19	288	22.8	248	2 Q44326	Q44326 agrobacteri
20	287.5	22.7	246	16 Q9KA03	Q9KA03 bacillus ha
21	287.5	22.7	257	16 Q92MR6	Q92MR6 rhizobium m
22	287.5	22.7	269	2 Q938D4	Q938D4 mycobacteri
23	286	22.6	247	17 Q9HVR8	Q9HVR8 halobacteri
24	285	22.5	247	16 Q8Y690	Q8Y690 listeria mo
25	283.5	22.4	256	16 Q982N5	Q982N5 rhizobium 1
26	282.5	22.3	245	16 Q8RDH9	Q8RDH9 thermomanae
27	282	22.3	260	5 Q15744	Q15744 dictyostell
28	282	22.3	284	10 Q50038	Q50038 nicotiana t
29	281.5	22.3	256	16 Q981H5	Q981H5 rhizobium 1
30	281.5	22.3	306	10 Q9LW34	Q9LW34 arabidopsis
31	281.5	22.3	306	10 Q941E4	Q941E4 arabidopsis
32	281	22.2	233	16 Q86732	Q86732 streptomyce
33	280.5	22.2	300	10 Q9LW35	Q9LW35 arabidopsis
34	280	22.1	255	16 Q8YL12	Q8YL12 anabaena sp
35	279.5	22.1	286	16 Q8UUT4	Q8UUT4 agrobacteri
36	278.5	22.0	244	16 Q55922	Q55922 synechocyst
37	278.5	22.0	291	4 Q9Y3A1	Q9Y3A1 homo sapien
38	278	22.0	246	16 Q9K636	Q9K636 bacillus ha
39	278	22.0	247	16 Q9PCN4	Q9PCN4 xylella fas
40	278	22.0	251	16 Q92PX8	Q92PX8 rhizobium m
41	278	22.0	255	2 Q07457	Q07457 rhodospendo
42	277	21.9	246	16 Q9A7P5	Q9A7P5 caulobacter
43	277	21.9	251	16 Q8YV10	Q8YV10 anabaena sp
44	277	21.9	256	16 Q92MR3	Q92MR3 rhizobium m
45	276	21.8	254	16 Q98BJ4	Q98BJ4 rhizobium 1

ALIGNMENTS

RESULT 1
ID Q9F4L7 PRELIMINARY; PRT; 250 AA.
AC Q9F4L7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ribitol dehydrogenase.
GN RTD.
OS Escherichia coli;
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C;
RA Lafayette P.R., Parrott W.A.;
RT "A non-antibiotic marker for amplification of plant transformation vectors in E. coli."
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
CC EMBL: AY005817; AAC01883.1; -
DR HSSP: O70351; 1E6W.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase.
SO SEQUENCE 250 AA; 26693 MW; 1787D4CEBFD6C6DE CRC64;

Query Match 100.0%; Score 1265; DB 2; Length 250;
Best local Similarity 100.0%; Pred. No. 66-91;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMHNSVPSMNTPLNGKVAALITGAASGIGLCQAKTLLDAGAKVYLIDEGSKLKIYVAELG 60
|||||
DB 1 MMHNSVPSMNTPLNGKVAALITGAASGIGLCQAKTLLDAGAKVYLIDREGDKLKIYVAELG 60

```
QY 61 ENAYALDLDLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAGSDPDVMDRVLTNLN 120
DB 61 ENAYALDLDLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAGSDPDVMDRVLTNLN 120
QY 121 INAFRCVAVLPFHIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVS 180
DB 121 INAFRCVAVLPFHIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVS 180
QY 181 QYGVAVGAVLPFPVYVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRD 240
DB 181 QYGVAVGAVLPFPVYVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRD 240
QY 241 LVILPGSVDL 250
DB 241 LVILPGSVDL 250

RESULT 2
O8YBCO PRELIMINARY: PRT: 242 AA.
ID O8YBCO;
AC O8YBCO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Ribitol 2-dehydrogenase (EC 1.1.1.56).
GN BMEI10980.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-16w / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; Pubmed=11756688;
RA Delvecchio V.G., Kapratel V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lyfidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Golsman E.,
RA Selskov E., Elzer P.H., Hagius S., O'Callaghan D., Telesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009731; AAL54222.1; -.
DR Interpro: IPR002198; ADH_short.
DR Pfam: PF00106; adn_short.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 242 AA; 25972 MW; E41DB241DB237D1 CRC64;

Query Match 66.68; Score 843; DB 16; Length 242;
Best Local Similarity 66.94; Pred. No. 4.6e-58;
Matches 162; Conservative 36; Mismatches 44; Indels 0; Gaps 0;

QY 9 MNPPLNGKVAITGAASGIGLOCAKTLDDAGAKVLLIDREDDKHKIYAEI GENAYALQL 68
DB 1 MAQDLNKRVAAVTGAASGIGLOCAKHKMIGATVFLVDRDENALAEKCEVLDGAKPLV 60
QY 69 DLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAGSDPDVMDRVLTNLNINAFRCV 128
DB 61 DLDLPASVAVMMPILEQAGLDIFHANAGATGCPVAGSDPDVMDRVLTNLNINAFRCV 120
QY 129 RAVLPFHIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVSQYGVAVGA 188
DB 121 HAVLPFHMERKGTGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVAPRGIRVG 180
QY 189 VLPFPVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRDVILPGSV 248
DB 181 VLPFPVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRDVILPGSV 240
QY 249 DL 250
DB 241 DL 242
```

```
RESULT 3
Q8U7X6 PRELIMINARY: PRT: 243 AA.
ID Q8U7X6;
AC Q8U7X6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Ribitol 2-dehydrogenase.
GN RDI OR ATU4323 OR AGR_L1076.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; Pubmed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kiteajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D., Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Shenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; Pubmed=11743194;
RA Goodner B., Hinkle G., Gätling Y., Miller N., Blanchard M.,
RA Gourollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollan C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009360; AAL45117.1; -.
DR EMBL: AE008252; AAK89116.1; -.
KW Complete proteome.
SQ SEQUENCE 243 AA; 26264 MW; AC363D61F1564DB2 CRC64;

Query Match 45.38; Score 573; DB 16; Length 243;
Best Local Similarity 46.18; Pred. No. 5.3e-37;
Matches 111; Conservative 52; Mismatches 78; Indels 0; Gaps 0;

QY 9 MNPPLNGKVAITGAASGIGLOCAKTLDDAGAKVLLIDREDDKHKIYAEI GENAYALQL 68
DB 1 MSELQKRIAVITTAASGIGLATTEALLEGATVYVMDNMKALNDLAKIGERAIQVY 60
QY 69 DLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAGSDPDVMDRVLTNLNINAFRCV 128
DB 61 NLDLADSCNAMIPEILGVDHIDILYCNAATYIGDLETTEPAIDKMLNINAVAMKNV 120
QY 129 RAVLPFHIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVSQYGVAVGA 188
DB 121 QAVVPFHMERKGTGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRROVAPRGIRV 180
QY 189 VLPFPVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRDVILPGSV 248
DB 181 VLPFPVATLDDMPKAKMEALANGSLMOPIEVASVLEFMTTRSKNVTVRDVILPGSV 240
QY 249 D 249
DB 241 D 241
```

```
RESULT 4
09LCV7 PRELIMINARY: PRT: 247. AA.
ID 09LCV7:
AC 09LCV7: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Clavalddehyde dehydrogenase.
GN CAD.
OS Streptomyces clavuligerus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1901;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL3585;
RA Aldoo K.A., Wong A., Alexander D.C., Rittammer R.A., Jensen S.E.;
RT "Cloning, sequencing and disruption of a gene from Streptomyces
RT clavuligerus involved in clavulanic acid biosynthesis.";
RL Gene 147:41-46(1994).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL3585;
RA Paraskar A.S., Aldoo K.A., Jensen S.E.;
RT "Molecular analysis of a beta-lactam resistance gene encoded within
RT the cephamycin gene cluster of Streptomyces clavuligerus.";
RL J. Bacteriol. 178:6266-6274(1996).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL3585;
RA Paraskar A.S., Aldoo K.A., Jensen S.E.;
RT "A pathway-specific transcriptional activator regulates late steps of
RT clavulanic acid biosynthesis in Streptomyces clavuligerus.";
RL Mol. Microbiol. 27:831-843(1998).
RN 14
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL3585;
RA Jensen S.E., Elder K.J., Aldoo K.A., Paraskar A.S.;
RT "Enzymes catalyzing the early steps of clavulanic acid biosynthesis
RT are encoded by two sets of paralogous genes in Streptomyces
RT clavuligerus.";
RL Antimicrob. Agents Chemother. 44:720-726(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: U87786; AAF86624.1; -.
DR HSSP: P14061; 1FDS.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
KW Oxidoreductase.
SQ SEQUENCE 247 AA; 26325 MW; 30A135E8CFE7E2E6D CRC64;

Query Match 28.9%; Score 366; DB 2; Length 247;
Best Local Similarity 37.3%; Pred. No. 7.5e-21;
Matches 91; Conservative 41; Mismatches 104; Indels 8; Gaps 3;

OY 9 MNTPLNGKVAATITGAAGTIGLQCAKTLTDAGAKVVLIDREGDKLHKIVAEFL--GENAYVA 65
DB 1 MPSALOGKVALITIGSSSGIGEARALALAEAGAAVAITAARVEKLRALGDELTLTAAGAKVHV 60
OY 66 LQDLFNNQOVNMLADIIELAGGLDIFHANAGAYIGPVAEGDPDVMRVNLININAAE 125
DB 61 LEIDVADRCGVDAVAASVTEALGLDILVNNAGIMILGPVEDADDTDMTRMIDTNILGLM 120
OY 126 RCYRAVLPHMIAQRSGDIITSSIAGVVPYIMEPIITASKFAVAQVHTTTRQVSOYQVR 185
DB 121 YMTRALPHLRSK-GTVVQMSIAGRVNNAAYQATKFGVNAFSETLROEVTGERGV 179
OY 186 VGAVLPGPVYAL---LDDWPKAKMEALANGSLMQPIEVAESVLFMTVRSKNVTVRDL 241
```

```
DB 180 VVIEFGTDTDELRGHTHTATKEWYEQRIQRLQADIAEAVRAVATPAHATVHEI 239
OY 242 VILP 245
DB 240 FIRP 243

RESULT 5
054197 PRELIMINARY: PRT: 247. AA.
ID 054197:
AC 054197: 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Clavulanate-9-aldehyde reductase.
GN CAR.
OS Streptomyces clavuligerus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1901;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 27064;
RA Perez-Redondo R., Rodriguez-Garcia A., Martin J.F., Liras P.;
RT "The car gene of Streptomyces clavuligerus, encoding a LysR-type
RT regulatory protein controlling clavulanic acid biosynthesis, is linked
RT to the clavulanate-9-aldehyde reductase (car) gene.";
RL Gene 211:311-321(1998).
RN 12
RP -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
RN (SDR) FAMILY.
CC CC
DR EMBL: AJ000671; CAA04227.1; -.
DR HSSP: P14061; 1FDY.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
KW Oxidoreductase.
SQ SEQUENCE 247 AA; 26627 MW; 251C1E2FCE6E75619 CRC64;

Query Match 27.0%; Score 342; DB 2; Length 247;
Best Local Similarity 35.7%; Pred. No. 5.6e-19;
Matches 87; Conservative 40; Mismatches 109; Indels 8; Gaps 3;

OY 9 MNTPLNGKVAATITGAAGTIGLQCAKTLTDAGAKVVLIDREGDKLHKIVAEFL--GENAYVA 65
DB 1 MPSALOGKVALITIGRELGHRATRALAPRGAATAARVEKLRALGDELTLTAAGAKVHV 60
OY 66 LQDLFNNQOVNMLADIIELAGGLDIFHANAGAYIGPVAEGDPDVMRVNLININAAE 125
DB 61 LEIDVADRCGVDAVAASVTEALGLDILVNNAGIMILGPVEDADDTDMTRMIDTNILGLM 120
OY 126 RCYRAVLPHMIAQRSGDIITSSIAGVVPYIMEPIITASKFAVAQVHTTTRQVSOYQVR 185
DB 121 YMTRALPHLRSK-GTVVQMSIAGRVNNAAYQATKFGVNAFSETLROEVTGERGV 179
OY 186 VGAVLPGPVYAL---LDDWPKAKMEALANGSLMQPIEVAESVLFMTVRSKNVTVRDL 241
DB 180 VVIEFGTDTDELRGHTHTATKEWYEQRIQRLQADIAEAVRAVATPAHATVHEI 239
OY 242 VILP 245
DB 240 FIRP 243

RESULT 6
0930M2 PRELIMINARY: PRT: 249. AA.
ID 0930M2:
AC 0930M2: 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Putative.
GN RA0173 OR SMA0320.
```

```

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymA (megaplasmid 1).
OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;
OC Rhizobiaceae: Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; Pubmed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Bariloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjel M., Hong A., Hultzer L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL: AE007211; MAK64831.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Plasmid: Hypothetical protein: Complete proteome.
SQ SEQUENCE 249 AA: 26553 MW: 0DFELF7597361E5E CRC64;

Query Match 26.3%; Score 333; DB 16; Length 249;
Best Local Similarity 34.4%; Pred. No. 2.8e-18;
Matches 84; Conservative 45; Mismatches 105; Indels 10; Gaps 4;

OY 11 TPLNGKVAATGAAGGIGLCCAKTLTDGAKVVLIDREGDKL-----KIYAEIGENAYAL 66
DB 2 TSLNGKIALVTGAGSSGIGATTAALAKLAGAKVGAARTRDLEDKKIEKKGSE-ALVI 60

OY 67 OLDENNOQVNMADLIEAGLIDIFHANAGAYIGGPAEGDPVDVRLNLINAFRCVR 126
DB 61 EMDVVDTSVEAGVKKLVDAVGSIDILVNAGLMLPSDIOFKKDEMQRWADVAVKGLLN 120

OY 127 CVRAVLEPHMAQNSGDIIFSSTAGVVPVIMEPIYTSKRAVQAQVHTTRQVSO-YGVR 185
DB 121 TTAIVLEIPIQMIKOHSGHFNNSSIGARKVFKGLSYCATKHAVTAFSGDLRMEVGOKKIGIR 180

OY 186 VGAVLPGPVYATLDDMP-----KAKMEBALANGSLMOPIEAEVLELPMVTRSKRVYTRDL 241
DB 181 VTCTGPFAVATELYDHTTDPGYRQGMDELATQMTFLQGEDIGDTIVRAQAPAHVDAEL 240

OY 242 VILP 245
DB 241 FVLTP 244

RESULT 7
OY9RF5 PRELIMINARY; PRT; 231 AA.
AC OY9RF5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein SAV2478.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
OC Bacteria: Firmicutes: Bacillus/Clostridium group: Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; Pubmed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mizutani-U I., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,

```

```

RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus";
RL Lancet 357:1225-1240(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: AP003365; BAB58640.1; -.
DR EMBL: AP003137; BAB43569.1; -.
DR HSSP: P50162; IAE1.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
KM Oxidoreductase: Hypothetical protein: Complete proteome.
SQ SEQUENCE 231 AA: 24578 MW: EED1BCE261E87778 CRC64;

Query Match 26.2%; Score 331; DB 16; Length 231;
Best Local Similarity 31.8%; Pred. No. 3.7e-17;
Matches 75; Conservative 54; Mismatches 99; Indels 8; Gaps 2;

OY 11 TPLNGKVAATGAAGGIGLCCAKTLTDGAKVVLIDREGDKLHKIVAEIGENAY-ALQLD 69
DB 2 TVLTDKIALVTGAGSSGIGATATLHLEGAKVVLAGRNKKDLQVAVNQLAQDSYKVVPTD 61

OY 70 LENNOQVNMADLIEAGLIDIFHANAGAYIGGPAEGDPVDVRLNLINAFRCVR 129
DB 62 VTNKEVDLMKKIAQGFEGGLDIVINSGOMLSKRTDYQVDEWSDMDVNIKCTLYTAQ 121

OY 130 AVLPHTIAQRSGDIIFSSTAGVVPVIMEPIYTSKRAVQAQVHTTRQVSOYGVRAV 189
DB 122 AALPTMLEOSSGHLINIAISGFEVTKSSITYSKFAVHTITQGLEKELAKTKVKTSTI 181

OY 190 LPGPVYATLDDMPKAKMEBALANGSLMOPIEAEVLELPMVTRSKRVYTRDVLTP 245
DB 182 SPGVVDTAITAAVNPSPRKK-----LDPODIAEAVLYALTKVHVNVEITVRP 230

RESULT 8
OY9RH22 PRELIMINARY; PRT; 248 AA.
AC OY9RH22;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Oxidoreductase.
OS Zymomonas mobilis.
OC Bacteria: Proteobacteria: alpha subdivision: Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2M4;
RA Lee J.S., Kang H.S.;
RT "Sequence analysis of 6563 cosmid clone of Zymomonas mobilis 2M4
RT containing rna operon.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL: AF088897; AAF18286.1; -.
DR HSSP: P14061; IEDS.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Oxidoreductase.
SQ SEQUENCE 248 AA: 26481 MW: 1C8335777C5E1EB1 CRC64;

Query Match 25.1%; Score 317; DB 2; Length 248;
Best Local Similarity 31.6%; Pred. No. 5e-17;
Matches 77; Conservative 53; Mismatches 106; Indels 8; Gaps 3;

OY 9 MNTPLNGKVAATGAAGGIGLCCAKTLTDGAKVVLIDREGDKL-----HKIVAEIGENAY 64
DB 1 MNQNRKIYVITGAGSGLGAETARHLSLDIGATVVLGARREERTATLANSTVAK-GGOAL 59

```

Db	288	ILADL	293
RESULT 10			
Q9UFM6	PRELIMINARY:	PRT:	325 AA.
AC Q9UFM6			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE Hypothetical 35.1 kDa protein.			
GN DKRP5660084.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX NCBI_TaxID=9606;			
RA [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=KIDNEY			
RX MEDLINE=21154917; PubMed=11230166;			
RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Glasl S.,			
RA Ansoirge W., Boecker M., Bloecher H., Bauersachs S., Blum H.,			
RA Laderer J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,			
RA Mewes H.-W., Oltewaldt B., Obermayer B., Tampe J., Heubner D.,			
RA Wambutt R., Korn B., Klein M., Postka A.;			
RT "Towards a catalog of Human Genes and Proteins: Sequencing and			
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."			
RL Genome Res. 11:422-435(2001).			
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES			
CC (SDR) FAMILY.			
CC EMBL: AL117567; CAB55997.1; -			
DR InterPro: IPR002198; ADH_short.			
DR Pfam: PF00106; adh_short; 1.			
DR PRINTS: PR00080; SDRFAMILY.			
KW Hypothetical protein; Oxidoreductase.			
SO SEQUENCE 325 AA; 35069 MW; F41956C54408BB43 CRC64;			
Query Match	24.0%;	Score 304;	DB 4; Length 325;
Best Local Similarity	33.3%;	Pred. NO. 7.3e-16;	
Matches 82; Conservative 38; Mismatches 108; Indels 18; Gaps			
QY 13 LNKRAATTGAASGIGLCQAKRTLDGAKRVYLIDREGDKLHKIVAEI-----GENA 63			
Db 50 LRNAVYVITGATSGIGKECAKVFYAGAKVLCGRNGALLEEIRLRTASHTKVTQTHKP 109			
QY 64 YALQDLLENNQOVDMLADIIELAGLDIEFNHAGAYIGAPAEQDPDWDVNLNINA 123			
Db 110 YLYFTFDLDSGAIYAAAHEILOCPCGYVDILVNNAGISTYGTIMDTFYVDVKRMETNYRG 169			
QY 124 AFRCYRAVLPHMIAPRSDDIIFTSSIAGVYVPIWEPIYTSKFAVQAVYHTTRROYSG 183			
Db 170 FVALTKALLPSMIKRRQGHIVASSIQGKMSIPFSAYAPRKHAFQDFDCLRAEMEOYE 229			
QY 184 VRVGVAVLEGPV-----VALLLDDMKR-AKMEALANGSLMOIEVAESVLFNV-TRSKRV 236			
Db 230 IEVYIISGPIHTNLSVNAITADGSRGCVMTTAAQG--RSPVEVAQDVLAIVGKKKKDV 287			
QY 237 TVRDLV 242			
Db 288 ILADL 293			
RESULT 11			
Q92NFB	PRELIMINARY:	PRT:	253 AA.
AC Q92NFB			
DT 01-DEC-2001 (TREMBLrel. 19, Created)			
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE Putative oxidoreductase protein.			
GN R03247 OR SMC01635.			
OC Rhizobium meliloti (Sinorhizobium meliloti).			
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			

```

OC Rhizobiaceae: Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gohit T., Goffeau A., Kahn D., Käss D., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591790; CAC46826.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 253 AA; 26818 MW; 0F2AAE2B89F4D8C3 CRC64;

Query Match 24.0%; Score 303.5; DB 16; Length 253;
Best Local Similarity 35.5%; Pred. No. 5.8e-16;
Matches 87; Conservative 43; Mismatches 100; Indels 15; Gaps 4;

OY 13 LNCVAITGAASGIGLQCAKTLIDAGAKVYLIDRE---GDKLHKIVAELEGNAVALQL 68
DB 3 LNNRIATVGTAGSGIGRAGAIAIRGCAHYVVDREVAAGDVFVAALTAG-GGSAEALAV 61

OY 69 DLFNNOGVNMLADIETELAGGLDIFHANAGAYIGGPAEGDPDWDVRLNINIAAFRCV 128
DB 62 DVYDDDLADGIADILYRHRGRIDILHNHAGOVAGDLEVEVAGFDSWNLNRAHFMMA 121

OY 129 RAVLPHMIAORSGDIIFTSSIAGVVPYIWEPIYASKFAVOAFVHTTRROYSGYEVGA 188
DB 122 RLVPSPKKKRGVYIVNTSSSGVLXDREMLAYTTTKHAYIAMTRQAGVAKYGVAVNA 181

OY 189 VLPGPVYVTFALLDWM-----PKAKMEFALANGSLMQPEVAESVLFNWT-RSKNVTY 238
DB 182 LCPGWPVTPNEPPIIDMGGRGREALIAYIRERVLGRMASVDEIAESILFLVSDRSYMTG 241

OY 239 RDLVI 243
DB 242 QILVV 246

RESULT 12
O9LA07 PRELIMINARY; PRT; 248 AA.
ID O9LA07:
AC O9LA07:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Hypothetical 26.4 kDa protein.
OS Paenibacillus popilliae (Bacillus popilliae).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=78057;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14706;
RX MEDLINE=20145393; PubMed=10681342;
RA Patel R., Piper K.E., Cockerill F.R., Steckelberg J.M., Yousten A.A.;
RT "The biopesticide Paenibacillus popilliae has a vancomycin resistance
RT gene cluster homologous to the enterococcal VanA vancomycin resistance
RT gene cluster.";
RL Antimicrob. Agents Chemother. 44:705-709(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
EMBL: AF155139; AAF36807.1; -.
DR HSSP: P14061; IEDS.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.

```

```

DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Hypothetical protein: Oxidoreductase.
SQ SEQUENCE 248 AA; 26432 MW; A122A9FD7799B10C CRC64;

Query Match 23.6%; Score 298; DB 2; Length 248;
Best Local Similarity 33.6%; Pred. No. 1.5e-15;
Matches 83; Conservative 43; Mismatches 105; Indels 16; Gaps 6;

OY 13 LNCVAITGAASGIGLQCAKTLIDAGAKVYLIDREDKHKIVAEI---GENAVALQL 68
DB 3 IENKVVITGASSIGGATKLLAEKAKVYLGRREHLVKLEELKNSGGAAYV-T 61

OY 69 DLFN---NQGVNMLADIETELAGGLDIFHANAGAYIGGPAEGDPDWDVRLNINIAAF 125
DB 62 DVVPPDSQQLVOLAKDTE--GGVDVIFLNAGLIMPNSPLELTKDEMSNVGVNKGVL 118

OY 126 RCVAVALPHMIAORSGDIIFTSSIAGVVPYIWEPIYASKFAVOAFVHTTRROYSG-- 183
DB 119 NGIAAVLPTFISQSGHITNWSVAGLKPAGCAVYATGVAANLMEVLRMESAQEGTN 178

OY 184 VRVGAVALPGPVVTFALLDWMFKAKME--ALANGSLMQPEVAESVLFNWT-RSKNVTY 240
DB 179 IRTATITPAINTLDTLTITDKNISSEGMTALYEDYGISPDVAVITAFALDQPEDTVNE 238

OY 241 LVILPGS 247
DB 239 FTIGPTS 245

RESULT 13
O9EX74 PRELIMINARY; PRT; 246 AA.
ID O9EX74:
AC O9EX74:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE SDR-like enzyme.
GN MHA.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC14;
RX MEDLINE=21091964; PubMed=11157238;
RA Van der Vliet-Berghmans C.J.B., van der Werf M.J.;
RT "Genetic and biochemical characterization of a novel monoterpene
RT epsilon-lactone hydrolase from Rhodococcus erythropolis DC14.";
RL Appl. Environ. Microbiol. 67:733-741(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
EMBL: AJ292535; CAC17805.1; -.
DR HSSP: PS0163; 2AEL.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Oxidoreductase.
SQ SEQUENCE 246 AA; 25058 MW; 6DECB136FE9AA34A CRC64;

Query Match 23.4%; Score 295.5; DB 2; Length 246;
Best Local Similarity 32.9%; Pred. No. 2.3e-15;
Matches 81; Conservative 42; Mismatches 110; Indels 13; Gaps 6;

OY 9 MNPPLNGVAITGAASGIGLQCAKTLIDAGAKVYLIDRGDKLHKIVAELEGNAVALQL 68
DB 1 MSYALEGKVAIVVTGGSGIGACVROLCALGASVVAIVDFNATLVAKFEFGRAVAVEV 60

OY 69 DLFNNOGVNMLADIETELAGGLDIFHANAGAYIGPV--AEGDP--DVMPRVNLINIAA 124
DB 61 DVARVEDAERKRVETAVAHFGGLDIAVNNAG--GVGPKASVGDGTGFEEMRRVLDVNLDA 118

```

ID	Accession	Gene	Protein	Score	DB	Length	Query Match
OY	125	FRCVAAVPHNIAORSGLIIFFTSIAGVVPVIMEPTTASKFAVOAFVHTTRRQVSGV	184	23.4%	Score 295.5	DB 1	Length 255
Db	119	FFCMRAELRRM--RKGGSVNVLASVMGAAVAGSSSYAASKALVGLTKTAALDYATACI	176	Best Local Similarity	30.0%	Pred. No. 2.5e-15	Mismatches 72; Conservative 57; Mismatches 104; Indels 7; Gaps 3
OY	185	RVGAVLPGRVVTALL----DDWPKAMEALANGSLMOPIEVAESVLEPMVTRSKNVTVAD	240				
Db	177	RVNAAAGAFVDPITLMAGRDPEW-LAAVAASHPLGRLAQDDEIASVVAFLASSAASFVIGA	235				
OY	241	LVILPG 246					
Db	236	FIPVDG 241					
RESULT 14							
ID	034187	PRELIMINARY	PRT	255	AA		
AC	034187						
DT	01-JAN-1998	(TREMBLrel. 05, Created)					
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)					
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)					
DE	Oxidoreductase.						
OS	Halobacterium volcanii (Haloferax volcanii).						
OG	Plasmid phv3.						
OC	Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.						
OX	NCBI_TaxID=2246;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=DS2;						
RA	Farahani R., Imbeault J.-C., St Jean A., Chan C.C.-Y., Allard G.,						
RA	Charlebois R.L.;						
RT	"Hereditary instability of the megaplasmid phv3, and filamentation in						
RT	the halobacterium Haloferax volcanii."						
RL	Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.						
CC	-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES						
CC	(SDR) FAMILY.						
DR	EMBL: U95375; AAB71809.1; -.						
DR	HSSP: 070351; 186M.						
DR	InterPro: IPR002198; ADH_short.						
DR	Pfam: PF00106; adh_short; 1.						
DR	PRINTS: PR00080; SDRFAMILY.						
DR	PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.						
KW	Oxidoreductase; Plasmid.						
SEQ	SEQUENCE 255 AA; 27018 MM; 7A768FF369BF24D CRC64;						
Query Match 23.4%; Score 295.5; DB 1; Length 255;							
Best Local Similarity 30.0%; Pred. No. 2.5e-15;							
Matches 72; Conservative 57; Mismatches 104; Indels 7; Gaps 3							
OY	13	LNKGAAATGGAAGSGIGCCAKTLLDGAAGVVLIDREGDKLHTVAVEL--GENAVAYLQD	69				
Db	10	LDGKAIVTIGAGSSGIGSATKSLASRGASVVAARREGLEELAAITIEDGDALVVPD	69				
OY	70	LENNQAVDNMLADIETLAGLIDIFHANAAGVYIGPYAEGDPDWMDFVNLNINAARCVR	129				
Db	70	VVDDIDIALVATLDEHGRIDILVNAGMLPLAHIGEADREFLTQTTIDVNLGLTLTH	129				
OY	130	AVLPHMIAQRSDIIFTSSIAGVVPVIMEPTTASKFAVOAFVHTTRRQVSGVAVGAV	189				
Db	130	AVAPPTMEQESHIVNLSVSGVRFLOANSNHNAAAGKMPKSDSLRLDVAAGIVVSI	189				
OY	190	LEGPVVTALLDDMPAKMEALAN--GSL--MQPIEVAESVLEPMVTRSKNVTVADVIAP	245				
Db	190	EGGAVDTLDDHIIPEEVQKNKVDYGTDALADAPEDIATITFFVYQPERVDINELYLR	249				
RESULT 15							
ID	09EGX3	PRELIMINARY	PRT	248	AA		
AC	09EGX3						
DT	01-MAR-2001	(TREMBLrel. 16, Created)					
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)					

Query	Match	Score	DB	Length	248
Best Local Similarity	32.1%	Pred. No. 4.4e-15;			
Matches	79;	Conservative	47;	Mismatches	106;
				Indels	14;
				Gaps	5;
QY	13 LNKVAAITGGAASGIGLCQAKTLLDAGAKVVIIDREGDKLHIV---AELGENVAYLDL	68			
DB	3 IKNKVIITGGAASGIGKKTALLAEKGAQLVAARVELEKIVTOTIRANSEALFA-KT	61			
QY	69 DLFNNQVVDNMLADIIELIAGGLIDIFHANAGAYIGGPVAEGDDVDVDRVLNINNAFCRV	128			
DB	62 DVTKREDNKKRLVELAIERYGKVDALIFLNAIGIMPNSPLSLAKEDDEWQMIDINIKGLVNGI	121			
QY	129 RAVLPHMTAAGSGDIIIFSSINGVYPVIMEPIYTAASKFVAQAFVHTTRQVQSYG--VRY	186			
DB	122 AAVLPSTFAQSGHITIAVSSVAAGLAAYPGAGVATKRWVRDLMEVLRHESAOEGNINFT	181			
QY	187 GAVLPGPVVTTALDDWPRAKMEALANGSL-----MQLIEVASVLFVNTRSKNVTVRDL	241			
DB	182 VTIYPAALNTELELTITTKETFEQGMT--SLYKQYGITPDRIASIYAVAIADGEDVDVNF	239			
QY	242 VILPGS 247				
DB	240 TVGPGS 245				

Search completed: March 13, 2003, 16:55:09
Job time : 18.9354 secs

RESULT 15		
09EXG3		
ID 09EXG3	PRELIMINARY;	PRT; 248 AA.
AC 09EG3:		
DT 01-MAR-2001	(TREMBLrel, 16, Created)	
DT 01-MAR-2001	(TREMBLrel, 16, Last sequence update)	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 ; Search time 43.7632 Seconds
(without alignments)
1625.927 Million cell updates/sec

Title: US-09-802-208B-4

Perfect score: 2811

Sequence: 1 MTIRKTVIGVDVSGSGSVRAG.....HHRREYAKYQLOHFAKLIRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2811	100.0	534	22	AAE09780
2	1172	41.7	551	23	AAE24140
3	1171	41.7	551	22	AAE40293
4	1005.5	35.8	548	22	ABR70032
5	672	23.9	246	22	AAE93914
6	553.5	19.7	554	22	AAU39145
7	448.5	16.0	560	19	AAW53950
8	435.5	15.5	593	22	AAU39179
9	434	15.4	252	22	AAE93394
10	340	12.1	518	23	ABP39224

11	330.5	11.8	521	22	AAE2039
12	330	11.7	164	22	AAE42079
13	320.5	11.4	505	23	ABP30486
14	320.5	11.4	518	23	ABP27721
15	317	11.3	519	23	ABE47468
16	309	11.0	519	23	ABE55539
17	306	10.9	484	22	AAU34790
18	306	10.9	498	23	ABE54582
19	304.5	10.8	494	22	AAE96826
20	303	10.8	501	22	AAU35273
21	302.5	10.8	503	22	AAU35487
22	298.5	10.6	497	23	ABE48654
23	298.5	10.5	496	22	AAE82146
24	293.5	10.4	497	23	ABP39566
25	289	10.3	517	22	AAU37244
26	288	10.2	509	22	AAU34300
27	281.5	10.0	508	22	ABP26834
28	279.5	9.9	509	22	AAE92932
29	279.5	9.9	509	22	AAE79335
30	277.5	9.9	502	23	ABP26833
31	276.5	9.8	497	20	AAW94843
32	276.5	9.8	502	22	AAU34832
33	276.5	9.8	502	22	AAU37838
34	273.5	9.7	502	22	AAU38067
35	272	9.7	538	22	ABE66928
36	269	9.6	498	17	AAU60271
37	265.5	9.4	526	17	AAW06440
38	265.5	9.4	501	21	AAV70727
39	264.5	9.4	496	20	AAW94309
40	264.5	9.4	496	22	AAU33825
41	264.5	9.4	498	22	AAU36742
42	262.5	9.3	501	22	AAU38195
43	261	9.3	501	23	ABE54851
44	259	9.2	514	23	ABP39502
45	255.5	9.1	499	22	AAE82195

ALIGNMENTS

RESULT 1	AAE09780	standard; Protein: 534 AA.
ID	AAE09780	
XX	XX	
AC	AAE09780;	
XX	XX	
DT	29-NOV-2001	(first entry)
DE	Escherichia coli strain C rfl operon encoding ribitol kinase.	
XX	XX	
KW	Positive selection system; metabolise: arabinol; ribitol; mannitol;	
KW	transgenic cell; marker gene: ribitol dehydrogenase; ribitol kinase;	
KW	ribitol transporter; rfl operon.	
XX	XX	
OS	Escherichia coli C.	
XX	XX	
PN	MO20016679-A2.	
XX	XX	
PD	13-SEP-2001.	
XX	XX	
PF	08-MAR-2001; 2001WO-US07474.	
XX	XX	
PR	08-MAR-2000; 2000US-0188291.	
PR	15-AUG-2000; 2000US-0255595.	
XX	XX	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.	
XX	XX	
PI	Parrott W, Lafayette P, Kane P;	
XX	XX	
DR	WPI: 2001-565596/63.	
DR	N-PSDB; AAD16811.	
XX	XX	
PT	Positively selecting transformed cells comprising selectable marker	

5. epidermidis ope
Human polypeptide
Streptococcus poly
Streptococcus poly
Streptococcus poly
Lactococcus lactis
E. coli cellular p
Lactococcus lactis
Putative P. abyssi
Enterococcus faeca
Haemophilus influe
Listeria monocylog
S. epidermidis ope
Staphylococcus epi
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Streptococcus poly
C glutamicum prote
Cornebacterium gl
Streptococcus poly
Heat-resistant gly
E. coli cellular p
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
Propionibacterium
Glycerol kinase.
Klebsiella oxytoca
Thermus flavus TE3
Staphylococcus aur
Staphylococcus aur
Salmonella typhi c
Lactococcus lactis
Staphylococcus epi
S. epidermidis ope

PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinol, ribitol which confer selective advantage on transformed
PS cells -
XX Claim 17; Page 37; 37pp: English.
CC The present invention relates to a positive selection system that
CC involves conferring to transferred cells the ability to metabolise
CC arabinol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C. ribitol kinase encoded by ribitol operon (rli operon). The
CC operon also encodes ribitol dehydrogenase and ribitol transporter.
XX
SQ Sequence 534 AA;
Query Match 100.0%; Score 2811; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 7, 6e-270;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 MTIRKTVIGVDVSGSVAGIFDLNGSLSHATEKITTTTRSGSRVDSOEIFMOAVCS 60
DB 1 MTIRKTVIGVDVSGSVAGIFDLNGSLSHATEKITTTTRSGSRVDSOEIFMOAVCS 60
OY 61 IRNALTLADVCAQSVAGIGFDATCSLVVLDKNQDPLPVSPBGDAKONIIYVMDHRAEQ 120
DB 61 IRNALTLADVCAQSVAGIGFDATCSLVVLDKNQDPLPVSPBGDAKONIIYVMDHRAEQ 120
OY 121 ERINATHHVPLVNYGKISPEMERPKILMKENKPEIYERAGOFDLDLFTIRATG 180
DB 121 ERINATHHVPLVNYGKISPEMERPKILMKENKPEIYERAGOFDLDLFTIRATG 180
OY 181 RSVCTVCKMTWLAHKNRMPDPRFTIGLAELADEDFIRIGHNHYVPTPCGNGLTAA 240
DB 181 RSVCTVCKMTWLAHKNRMPDPRFTIGLAELADEDFIRIGHNHYVPTPCGNGLTAA 240
OY 241 AEMGLPFGTPYAVGLIDAHAGIGTVGEGALNNLAVFEGTSSCTMASTTSPSFV 300
DB 241 AEMGLPFGTPYAVGLIDAHAGIGTVGEGALNNLAVFEGTSSCTMASTTSPSFV 300
OY 301 GPYISAMVPGMLVEGGOSAGAAIDOLDHPHVAEAREMAORVNOPLPWLADRI 360
DB 301 GPYISAMVPGMLVEGGOSAGAAIDOLDHPHVAEAREMAORVNOPLPWLADRI 360
OY 361 TAQPSDAVALAKGLHVVPEFLGNRAPPARAVICGLGEMERLDNLALYIGLCIG 420
DB 361 TAQPSDAVALAKGLHVVPEFLGNRAPPARAVICGLGEMERLDNLALYIGLCIG 420
OY 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROILDTCIPYITTOCCPEVLGSA 480
DB 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROILDTCIPYITTOCCPEVLGSA 480
OY 481 GAVAGNIAPSVGEAMQOFTHYDKYYPQERYOSLHRRYEAYKOLQHTAKLLRD 534
DB 481 GAVAGNIAPSVGEAMQOFTHYDKYYPQERYOSLHRRYEAYKOLQHTAKLLRD 534
RESULT 2
ID AAE24140 standard; Protein; 551 AA.
XX AAE24140:
AC AAE24140:
XX
DT 23-SEP-2002 (first entry)
XX

DE Human kinase (PKIN)-11 protein.
XX
XX Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;
KW acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
KW asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;
KW development; hepatitis; cardiovascular; hypertension; drug screening;
KW myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;
KW fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;
KW hypercholesterolaemia; obesity; gene therapy; cytostatic; anti-HIV;
KW neuroprotective; hepatotropic; hypotensive; cardiatic; nephrotropic;
KW hyperlipidaemia; enzyme.
XX
OS Homo sapiens.
XX
PN WO200233099-A2.
XX
XX 25-APR-2002.
XX
PF 20-OCT-2001; 2001WO-US47728.
XX
XX
PR 20-OCT-2000; 2000US-242410P.
PR 27-OCT-2000; 2000US-244068P.
PR 03-NOV-2000; 2000US-245708P.
PR 09-NOV-2000; 2000US-247672P.
PR 16-NOV-2000; 2000US-249565P.
PR 22-NOV-2000; 2000US-252730P.
PR 01-DEC-2000; 2000US-250807P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvizu C;
PI Yao WG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;
PI Lal PG, Recipon SA, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM;
PI Thangavelu K, Khan FA, Ison CH;
XX
DR WPI: 2002-454603/48.
DR N-PSDB: AAD38854.
XX
PT New human kinase polypeptide, for diagnosing, preventing and treating
PT cancer, immune system disorders, growth and development disorders,
PT cardiovascular disorders and lipid disorders -
XX
PS Claim 1; Page 160-161; 210pp: English.
XX
XX The invention relates human kinases (PKIN) and their corresponding
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
CC treating and preventing cancer, an immune system disorder (e.g.,
CC acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,
CC asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders
CC affecting growth and development (e.g., arteriosclerosis, cirrhosis,
CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial
CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty
CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,
CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
CC condition or a disease associated with the expression of PKIN in a
CC biological sample. A composition comprising PKIN or an agonist or
CC antagonist of PKIN is useful for treating a disease or condition
CC associated with decreased or increased expression of functional PKIN.
CC PKIN is useful in a number of drug screening techniques and to analyse
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
CC knockin humanised animals or transgenic animals to model human diseases,
CC and in somatic or germ-line gene therapy. The present sequence is human
CC PKIN protein.
XX
SQ Sequence 551 AA;
Query Match 41.7%; Score 1172; DB 23; Length 551;
Best Local Similarity 45.0%; Pred. No. 4, 3e-107;
Matches 242; Conservative 96; Mismatches 182; Indels 18; Gaps 8;
OY 8 IGVDSGVSAGIFDLNGSLSHATEKITTTTRSGSRVDSOEIFMOAVCSIRNALTL 67

```
Db      13 VGVAVGTGSGVRALVDQSGVLLAFADQPIKNWEPGFNNHSESSSEDIWAACCVTKKRYVG 72
Oy      68 ADVCAQSVAGIGFDATCSLVVLDKNGDPLVPSPGDAKONIYMMHRRATEQAERINATH 127
Db      73 IDL--NQIRGLGFDAATCSLVVLDKOFHPLRVNQGDSHRRVIMWLDHRAVSQNRINETK 130
Oy      128 HPLVNVGKISPEMETPKILMLKENMPET--YERAGOFEDLADFLWRATGDLARSVCY 186
Db      131 HSLVQVGVGVSMVEKMAPKILMLKENLREICMDKACHFFDLPFLSKKATGYARSLCSL 190
Oy      187 TCKWTWLAHENRMDPDYFRTIGLAELADEDFIRIGHIIVSPGRCNGLTQAQAAEMGILL 246
Db      191 VCKWTYSA--EKGMDDSFWMKMGIEDEFVADNYSKIGNQVLPFGASLNGLTPEAARDGLIL 249
Oy      247 PGTPVAVGLIDAHAGIGTVG--VEGGAL-----NNLAUYFGTSSCTMASTSPSFV 296
Db      250 PGIAVAASLIDAHAGLGVIGADVGRHGLICEQPYTSRLAVICGTSSCHMIGISKDPIFY 309
Oy      297 PGVWGPYYSAMVPGLMLVEGSGSAGAAIDQLDFHPAVEAREMAQRVNPPLVWLADR 356
Db      310 PGVWGPYFSAMVPGFMLNEGGSQVTKLIDHMOGHAAPFELQVKATARCQSIYAYLNH 369
Oy      357 I-LEKTAQPSDAVALAKGLHNVPEFLGNRPADPHARAVYICGLMERDIDLALYIAG 415
Db      370 LDIKKAQF--VGFLLTVDLHWMPDFHGNRSPLADLTCLKGWTGKLKSQDLDDLAIILYAT 427
Oy      416 LCGIGVGLROIIDAQAGVSNKIYISGAGOHPLVROLADPGCIPVTTGCCPEVLL 475
Db      428 VOAILGSTRFITEAMEAGHSISTFLCGGLSNPLFVONHADITGMPVVLSDQEVESVLV 487
Oy      476 GSAILGAVAGNIAPSVGEAMQCFTHYDKYUYPOERYQSLHRRYEAV--KOLQHTAKLL 532
Db      488 GAAVLGACASGDFASVGEAMAKMSKVYFRLQDKKYDKYQVFLKLVENQKETL 545
```

```
RESULT 3
AAM40293
ID AAM40293 standard; Protein: 551 AA.
```

```
AC AAM40293;
```

```
DT 22-OCT-2001 (first entry)
```

```
DE Human polypeptide SEQ ID NO 3438.
```

```
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.
```

```
OS Homo sapiens.
```

```
XX PN WO200153312-A1.
```

```
XX PD 26-JUL-2001.
```

```
XX PF 26-DEC-2000; 2000WO-US34263.
```

```
XX PR 21-JAN-2000; 2000US-0488725.
```

```
XX PR 25-APR-2000; 2000US-0552317.
```

```
XX PR 09-JUL-2000; 2000US-0598042.
```

```
XX PR 19-JUL-2000; 2000US-0620312.
```

```
XX PR 03-AUG-2000; 2000US-0653450.
```

```
XX PR 14-SEP-2000; 2000US-0662191.
```

```
XX PR 19-OCT-2000; 2000US-0693036.
```

```
XX PR 29-NOV-2000; 2000US-0727344.
```

```
XX PA (HYSE-) HYSEQ INC.
```

```
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
```

```
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimaane RT;
XX N-PSDB; AA159449.
```

```
XX WPI: 2001-442253/47.
```

```
XX Novel nucleic acids and polypeptides, useful for treating disorders
```

```
XX such as central nervous system injuries -
```

```
XX Example 5; SEQ ID NO 3438; 10078bp; English.
```

```
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
```

```
XX Sequence 551 AA:
```

```
XX Query Match 41.7%; Score 1171; DB 22; Length 551;
```

```
XX Best Local Similarity 45.0%; Pred. No. 5,5e-107;
```

```
XX Matches 242; Conservative 96; Mismatches 182; Indels 18; Gaps 8;
```

```
Oy 8 IGVAVGSGVRAGIFDLNGSLSHATEKITTTTRSGRVSQSSQETMQACSCIRNATLT 67
```

```
Db 13 VGVAVGTGSGVRALVDQSGVLLAFADQPIKNWEPGFNNHSESSSEDIWAACCVTKKRYVG 72
```

```
Oy 68 ADVCAQSVAGIGFDATCSLVVLDKNGDPLVPSPGDAKONIYMMHRRATEQAERINATH 127
```

```
Db 73 IDL--NQIRGLGFDAATCSLVVLDKOFHPLRVNQGDSHRRVIMWLDHRAVSQNRINETK 130
```

```
Oy 128 HPLVNVGKISPEMETPKILMLKENMPET--YERAGOFEDLADFLWRATGDLARSVCY 186
```

```
Db 131 HSLVQVGVGVSMVEKMAPKILMLKENLREICMDKACHFFDLPFLSKKATGYARSLCSL 190
```

```
Oy 187 TCKWTWLAHENRMDPDYFRTIGLAELADEDFIRIGHIIVSPGRCNGLTQAQAAEMGILL 246
```

```
Db 191 VCKWTYSA--EKGMDDSFWMKMGIEDEFVADNYSKIGNQVLPFGASLNGLTPEAARDGLIL 249
```

```
Oy 247 PGTPVAVGLIDAHAGIGTVG--VEGGAL-----NNLAUYFGTSSCTMASTSPSFV 296
```

```
Db 250 PGIAVAASLIDAHAGLGVIGADVGRHGLICEQPYTSRLAVICGTSSCHMIGISKDPIFY 309
```

```
Oy 297 PGVWGPYYSAMVPGLMLVEGSGSAGAAIDQLDFHPAVEAREMAQRVNPPLVWLADR 356
```

```
Db 310 PGVWGPYFSAMVPGFMLNEGGSQVTKLIDHMOGHAAPFELQVKATARCQSIYAYLNH 369
```

```
Oy 357 I-LEKTAQPSDAVALAKGLHNVPEFLGNRPADPHARAVYICGLMERDIDLALYIAG 415
```

```
Db 370 LDIKKAQF--VGFLLTVDLHWMPDFHGNRSPLADLTCLKGWTGKLKSQDLDDLAIILYAT 427
```

```
Oy 416 LCGIGVGLROIIDAQAGVSNKIYISGAGOHPLVROLADPGCIPVTTGCCPEVLL 475
```

```
Db 428 VOAILGSTRFITEAMEAGHSISTFLCGGLSNPLFVONHADITGMPVVLSDQEVESVLV 487
```

```
Oy 476 GSAILGAVAGNIAPSVGEAMQCFTHYDKYUYPOERYQSLHRRYEAV--KOLQHTAKLL 532
```

```
Db 488 GAAVLGACASGDFASVGEAMAKMSKVYFRLQDKKYDKYQVFLKLVENQKETL 545
```

```
RESULT 4
ABB70032
```


XX AAW53950;
 AC 17-AUG-1998 (first entry)
 DT
 XX
 DE Bacillus subtilis L-ribulokinase.
 XX
 KM Promoter: ara operon; heterologous gene expression; L-arabinose;
 KM arab; L-ribulokinase.
 XX
 OS Bacillus subtilis.
 XX
 PN W09811231-A1.
 PD 19-MAR-1998.
 XX
 PF 10-SEP-1997; 97MO-US16097.
 XX
 PR 10-SEP-1996; 96US-0031077.
 XX
 PA (UYRQ) UNIV ROCKEFELLER.
 PI De Lencastre H, De Saenoguelra I;
 XX
 PI WPI: 1998-207392/18.
 DR N-PSDB; AAV25568.
 XX
 PT L-arabinose-utilising operon promoter - useful for regulation of
 PT heterologous protein expression in prokaryotic host cell
 XX
 PS Example 1; Fig 2A-B; 83pp: English.
 XX
 CC This polypeptide comprises the L-ribulokinase of Bacillus
 CC subtilis, encoded by the arab gene of the ara operon (see AAV25568).
 CC This enzyme is involved in the utilisation of L-arabinose by
 CC B. subtilis. The invention relates to the strong promoter (see
 CC AAV25567) of the ara operon, which is induced by L-arabinose and
 CC repressed by glucose, and its use in regulating the expression of
 CC heterologous genes in prokaryotic hosts.
 XX
 SQ Sequence 560 AA;
 Query Match 16.0%; Score 448.5; DB 19; Length 560;
 Best Local Similarity 27.1%; Pred. No 3e-35;
 Matches 133; Conservative 95; Mismatches 237; Indels 79; Gaps 20;
 QY 8 IGVNDSGSGSVRAGIFDL-NGSLSHATEKI-----TTTRSGSRVE-----OSSQRIWQ 55
 DB 5 IGVDFGLSGRAVLVHVQTEELAAAVKEYRHAVIDTVLPRTGQKLRDMALQHPADYLE 64
 QY 56 AVSCIRNATLADYCAQSVAGTGFQAT-CSLVVLDKNGDPLVSPRPGDAKON--ITVMM 112
 DB 65 VLETTTIPSLLEQGTGVDPKDIDIGIDFTACTITPIDSSGQPLCMLPVEEPSPHYVXLMK 124
 QY 113 DHRATEQAEKIN-----ATHNPLVNVGSKTSPMEPRKILMKENMEITEIRAGQFFDLA 168
 DB 125 HHAQAQKADLNLQIAEEGGAFLQRTGGKTSSEMMIPKMQIAEEAHITTEADRIIEAA 184
 QY 169 DFLTWRAQGLDARSVCVTCTKMTWLAHENRMDPDYFTIGLA-ELADEDFIRIGHNIVSP 227
 DB 185 DWLVYQDGLSKRSNCTAGKAKMSEKAGPSDDEFKELMPSKMTTKD--KLSGSHSV 242
 QY 228 GTPCGNGLTAQAAAMGLPRTPAVAGLIDAHAGIGITGV-ESGALNINLAAYFGTSSCT 266
 DB 243 GERAGS-LTFEKMAKGLILPSTAVAVANVAHV-SVPAVITEPG---KMLMIMGTSTCH 297
 QY 287 MASTSPSPRPYGVWPRYSAMVGLMVEGGSAGALDQLD--FHRAV-EEARE--- 340
 DB 298 VLLGEEHVHYVPGMGVVDNGLIPGYAGYEGQSCVGHDFWFKTCVPAYOEAEKNI 357
 QY 341 -----MAORVNOPLPVWMLDRILEKTAQPSDVAALAKGLHVPEFLGNRPFDPHRAV 395
 DB 358 GVHELLSEKANHQAFC-----ESGLLALDMMWNGNRSTLVADDLTGM 398

QY 396 ICGMERDLDLNLALYIAGLCGIGYGLRQILDQAQGVSKNIVISGG-AGQHPLVRO 454
 DB 399 LIGMTL---LTKPREIVRALVEATVGRMIIEFKESGVIEELFAAGIAEKNPVMO 455
 QY 455 ILADTCGPVYTTQCCERVLVLSAIIICAVAG-----NAPSVEAMQOFTIVDK 503
 DB 456 IYADVNTMMDIKISGSPAPALGSAIFGALAGKEKGGYDIDIKAAAMNGKIK-----DI 509
 QY 504 YVYPOERYOSLHRRYEAAYKOLQ 527
 DB 510 TYTPNAENAAVYKELVAYEKELVH 533
 RESULT 8
 ID AAV39179 standard; Protein: 593 AA.
 XX
 AC AAV39179;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #75.
 XX
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN W0200181581-A2.
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001MO-US12665.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Shelly YW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'atsosonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI: 2001-616774/71.
 DB N-PSDB; AAS59506.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Example 1; SEQ ID NO 374; 1069pp: English.
 XX
 CC Sequences AAV39105-AV68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed

DT 25-JUL-2002 (first entry)
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4069.
DE Staphylococcus epidermidis: open reading frame; ORF; bacterial infection;
XX Staphylococcus epidermidis: gene therapy.
KM antibacterial.
XX Staphylococcus epidermidis.
OS US6380370-B1.
XX US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-0134001.
XX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PA Doucette-Stamm LA, Bush D;
PI WPI: 2002-381255/41.
DR N-PSDB: ABN91769.
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX Disclosure; SEQ ID 4069; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX Sequence 518 AA;
SQ
Query Match 12.1%; Score 340; DB 23; Length 518;
Best Local Similarity 25.4%; Pred. No. 1.6e-24;
Matches 137; Conservative 85; Mismatches 236; Indels 82; Gaps 19;
QY 25 NGSLSHATEKITTRRGSRVDSOSRINQAVCSIRNALTLDYCAQSVAGIGFDATC 84
DB 18 NGTYMNSLYDK---PLPGNFFLQADYDQLLEGVQVLEDSKVNKNNDVVGIGVDT 73
QY 85 SLVY-LDKNGDPLVSPEDAKON---IIVMDHRAEQAEKRI--NATHHPLVLYVCGK 137
DB 74 STIFLDEQFRL--HREDKATNPNAVVKMKHHGADENAYMOMSKNNKMLDYISS 131
QY 138 ISPEMERKILMKENMEIYERAGQFADLFTLWRAATGDLANSVCTVTKWTWLAHEN 197
DB 132 VNSEMMIPKILEVHKEAPEILRRARYIEMAGDYITSLTNSINSNCIGKFGW-DNEA 190
QY 198 RMDDEYRTITGLABDEDFIRIGHN-----IVSPGTCGGLTRQAAMENGLRPTVA 252
DB 191 GFNDVDFHSV-----DPLDKIVKEKCEAPDISGESAGR-LCKDYQOILSDQDVVS 243
QY 253 VGLIDAHAGIGTGVVEGALNLAUVFGTSSCTMASTTSFVPGVWGPYYSAMVPLTW 312
DB 244 PFIIIDAHSGVGVGAIEG---EFTAVIGTICHLMDSRVSPISITGSVKNAIIPELY 300
QY 313 LVESGQSNAG-----AAIDQLDFH-PAVEARMAQRVNOPLPFWMLADRIE 359
DB 301 AVEAGQPAVGDLEFYSKNQAPKRIIVQANENHMPVNLVLEBLASHI-----RIEE 350
QY 360 KTAQPSDAVALAKGLHVPERL-GNRAPFADPHARAVYCGIGMERDNLALYIAGICG 418

DB 351 QHV-----VYLDWLNGRSILSNHLTGSIFGLTLQPYEMIRAYE---A 394
QY 419 IGYGLROIIDAQTACGVYSKNIVISGAGQ--HPIVROLADPCGIPVTTGCCCEVLGGS 477
DB 335 TAFGTRKIMKQFEDNHHIVHYHYVASGIGIPKSKLVEIYANLVKRRVYIDSSNASALGA 454
QY 478 AILGAVAGNIAPSVGEA---MQQFTHVDRKYYPQERYSLHRRYEAAYKOLQHTAKLLRD 534
DB 455 AMLGANVGNNAVSTLKEALSKMQ-----PIAYIGEPLEIKVQAVKPYHKYCELDH 505
RESULT 11
AAG82039
ID AAG82039 standard; Protein; 521 AA.
XX AAG82039;
AC AAG82039;
XX 03-SEP-2001 (first entry)
DT 5. epidermidis open reading frame protein sequence SEQ ID NO:1172.
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
DE vaccination; endocarditis.
XX Staphylococcus epidermidis.
OS WO200134809-A2.
XX WO200134809-A2.
PD 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US30782.
PF 09-NOV-1999; 99US-0164258.
PR (GLAX) GLAXO GROUP LTD.
XX kimerly WJ;
XX WPI: 2001-316495/33.
DR N-PSDB: AAH52889.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 337; 2188pp; English.
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the amplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454, so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 521 AA;
SQ
Query Match 11.8%; Score 330.5; DB 22; Length 521;
Best Local Similarity 25.1%; Pred. No. 1.4e-23;
Matches 135; Conservative 86; Mismatches 252; Indels 65; Gaps 17;


```

OY 25 NSLSLHATEKITTTTRRSRVE-----OSSOEIOWAVCSICRNALLTLADCAQSV 75
DB 8 NCHITSRYEDYANGTYMNSLYDKPLPENYFLIONADDTYLIIEOGVQFVLEDSKVKNDV 67
OY 76 AGIGDATCSLVV-LDKNGDPLVPSPEGDAKON---IIVMDHRAEQAEIRI--NATHH 128
DB 68 VGIGVDFSTSTIIFLEDEQEP--HRHEDLKTNPNAVYVLMKHKHGAODEANVYIQGSKNK 125
OY 129 PVLNVYGGISPEMETPKILMKENPEIYERAGOFFDLADFLTPRATGDLARSVCTVTC 188
DB 126 NMLDYVGGSSVNSEMMIPIKILEVKNHAPILRRARYIMEAGDYTSTILTSNINRSNCIGIF 185
OY 189 KMTVLHNERMWDYFRTTIGLAELADEDFLRIGHN-----IYEPGTPCNGSLTAQAAM 243
DB 186 KGW-DNEAGFNDFPHSV-----DPDLPKYKECEAPITISGSAGR-LCKDYQQLW 237
OY 244 GLLPGTPNAVGLIDAHAGIGTIVGEGALNMLNAVYEGTSSTMASTTSPSFVPGWGPY 303
DB 238 GLSQYVQVSPFITIDAHSGVIGVATGAEAG---EFTAVIGTSTCHMLDSRQVPISITGSV 294
OY 304 YSAMVNGVLVLEGGSAAAGAIQDLDFHR--AVEBARMAOVNOLPLVWGLADRLLEKT 361
DB 295 KNAIIPGLAYEAGOPAVGDLFEYSKQAPKNIIVDQANEHMHV-----LNYL 342
OY 362 AOPSDAVALAKGLHVPPEFL-GNRAPFADPHARAVICGLMERDLDNLALYTAGLCIG 420
DB 343 ELASIRIEEONVYVLDMLNGNRSLNSHLTGSIFGLTLQTPYEMIRAYIE---ATA 399
OY 421 YGLRQILDAOTAGVSVKNIVISGAGQ--HPLVRQILADTCGIPVITTOCCPEVLLGSAI 479
DB 400 FGKFLIMKQFEDNHIPVHFTVYASGAPIQKSKILLVEIYANVLRKRVYIDSSNASALGAM 459
OY 480 LGAVNAINPVSVEA---MQQFTHVDKYYPQGRYOSLHNRREAKQOLHTKLLRD 534
DB 460 LGANVNAVSTLEKALSMKQ-----PIAYIOEPLOKQVAKPKYHKXCELMD 508

RESULT 12
AAM42079
ID AAM42079 standard; Protein; 164 AA.
AC AAM42079;
XX
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 7010.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
XX
XX MO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA

```

```

XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dymanc RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI61235.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 7010; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 164 AA:
XX
XX
XX Query Match 11.7%; Score 330; DB 22; Length 164;
XX Best local Similarity 50.4%; Pred. No. 2.6e-24;
XX Matches 65; Conservative 20; Mismatches 42; Indels 2; Gaps 1;

OY 8 IGVVSGSVYRACIGFLNGLSHATEKITTTTRRSRVEOSSOEIOWAVCSICRNALLTL 67
DB 38 VGVDVGGSVRYALALVDQSGVLLAFADQPIKNMPEQFNHHEQSSBEDIACCVYTKVYQG 97
OY 68 ADVCAQSVAGIGDATCSLVVLDKNGDPLVPSPEGDAKONIIVMDHRAEQAEIRNATH 127
DB 98 IDL--NQIRGLGFDATCSLVVLDKOFHPLPVNOGBDSHRNVIWMLDHRASQVNRINERK 155
OY 128 HPVLVYVGG 136
DB 156 HSYLQYVGG 164

RESULT 13
ABP30486
ID ABP30486 standard; Protein; 505 AA.
AC ABP30486;
XX
XX 02-JUL-2002 (first entry)
DE Streptococcus polypeptide SEQ ID NO 10148.
XX
XX Streptococcus agalactiae.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX

```



```

OY      8 IGVVSGSVRAGIFDLNGLSLSHAT-EKITTTRSGSRVSOSEIOMAVCSIRNALR 66
DB      18 LSLDYGCTMTKALIFPKLHQIAVSSFEIKNTQSGHR-QVNLVTWNAITTAIEVIO 76
OY      67 LADVCAOSVAGIGFDATC-----SLVYLDKNGDPLFVSPEDAKONIYVMDHRAEQAE 121
DB      77 ISKLSPEQISAV-----ACIGHGKGLYLDNKLLEPL-----EGGILSTDRKADLQ 123
OY      122 RINATHHPLVNYGKISPEMETPKIL-WLKEMPEIYERAGOFPLDPLTRARAGDLA 180
DB      124 YFESKLDNITWELTROHIFPS-OSPVILRLMLKDYQPEFYKSIGAVLSAKPIRYKLTKYQ 182
OY      181 RSVCTVCKMTWLAHENRW-----DP---DYF---RTI---GLAEADE-DFIRIGHNI 224
DB      183 QEVGD-----ASGNHMINFOTGYDDAIDDFGIRLEINSLPELIDBADLV----- 228
OY      225 VSPGTPCGNLTAQAAAEKMLPGRTPVAVGLIDAHAGIGTGVEGGALNNLAVFGTSS 284
DB      229 --PG-----GISQAARKETGLVEGTPVVGGLFDIDACALSGVLESD-----TFEVI 274
OY      285 CTMASTTSFVP-----GWMGPYYSAMVPGMLVLEGGSAAGAALDOLD--FHPAVEE 337
DB      275 GTWNIWTPYSLKPAKQDSGLMTSFP---DRVLLERASSPTSGAGNLFMLKMLHDEIDN 331
OY      338 AREMAORVNOPLPWLADRILEKTAOPSDAVALAKGLHVPEFLGNRAPDPHARAVIC 397
DB      332 AKSSGSIYDNLEPL-----THTDATHHGLIFPEPLXGSN---TSODASACFF 377
OY      398 GIGMERDLDNL-ALYIAGIGIGYGLRQ-ILDAQTAQGVASKNIYVSGAGOHPLVRQI 455
DB      378 GLTTKSTKSQMIKRAVE-----GIAFAKHQITDLIKSRGSPVKIIRPSGATNSPAMQM 433
OY      456 LAFPCGIPVTTGCCPVLIGSAIILGAVAGNIAPSVGEAMQOPTHDKYKYPQERTGSLH 515
DB      434 FSOILNFPLETTVEGTGLGGAILARHALD-KISLKEAVODMVRKAIYKPOLSEVKG 492
OY      516 HRRYEAVKQLOHT 528
DB      493 KKKYHAYOKLLET 505

RESULT 15
ID      ABB47468
AC      ABB47468 standard; Protein; 504 AA.
XX
XX      ABB47468;
XX
XX      05-FEB-2002 (first entry)
XX
XX      Listeria monocytogenes protein #172.
XX
XX      Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX      vitamin B12; bacterial infection; disease.
XX
XX      Listeria monocytogenes.
XX
XX      MO200177335-A2.
XX
XX      18-OCT-2001.
XX
XX      11-APR-2001; 2001MO-FR01118.
XX
XX      11-APR-2000; 2000FR-0004629.
XX
XX      (INSP ) INST PASTEUR.
XX
XX      Buchliesser C, Frangoul L, Couve E, Rusnlok C, Psahi H, Dehoux P;
PI      Dussarquet O, Chetouani F, Nedjati H, Glaser P, Kunst F, Cossart P;
PI      Daniels J, Goebel W, Krefte J, Kuhn M, Ng E, Vazquez-Boland JA;
PI      Dominguez-Bernal G, Garrido-Garcia P, Tlierrez-Martinez A, Amend A;
PI      Chakraborty T, Domann E, Hain T, Berche P, Chardit A, Durant L;
PI      Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

```

```

PI      Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI      Rose M, Voss H;
XX
XX      WPI; 2002-010914/01;
XX
XX      genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX      and prevention of Listeria and related bacterial infections, and
XX      related polypeptides;
XX
XX      Claim 6; SEQ ID No 173; 192pp; French.
XX
XX      The present invention relates to the genome sequence of Listeria
XX      monocytogenes EGD-e (see AB03041). The genome sequence and fragments of
XX      it are useful for selecting probes and primers for detecting genes in L.
XX      monocytogenes and related organisms, and for studying genetic
XX      polymorphisms and other genomes. The present sequence is a protein
XX      encoded by the genome sequence of the present invention. Proteins
XX      expressed from the genome sequence are useful for raising specific
XX      antibodies, identification of L. monocytogenes and related organisms, and
XX      for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX      B12. The genome sequence and proteins encoded by it are also useful for
XX      selecting compounds that regulate gene expression and cell replication
XX      and modulate L. monocytogenes-related diseases. In addition, the genome
XX      sequence and proteins encoded by it are useful in pharmaceutical and
XX      vaccines compositions for the treatment or prevention of infections by L.
XX      monocytogenes and related organisms.
XX      Note: The sequence and data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX      Sequence 504 AA:
XX
XX      Query Match 11.3%; Score 317; DB 23; Length 504;
XX      Best Local Similarity 23.7%; Pred. No. 2.9e-22;
XX      Matches 132; Conservative 98; Mismatches 240; Indels 88; Gaps 22;
OY      1 MTTTKVIGVGVSGSVRAGIFDLNGLSL-SHATEKITTRSGSRVSOSEIOMAVCS 59
DB      1 LTKSPYIMGVDIGSSRKAVLFFNORGVEIFRQATHYELINDRG-KAESPTLEIDVLFT 59
OY      60 CIRNALLADVCAOSVAGIGF-DATGSLVYLDKNGDPLFVSPEDAKONIYVMDHRAE 118
DB      60 SIQ-AVAKNINTELAGISFSSAMSLIMVSGELL-----TECITADNGSSE 108
OY      119 QAEIRINATHHPLVNY--VGKISPEMETPKILWLKEMPEIYERAGOFPLDPLTRARAT 176
DB      109 TLEKVKRDNITLFQLYEATGPIPIHMSPPAKICWLKEKPTLFKTEKTFVDIKSYILRL 168
OY      177 G--DLARSVCTVCKMTWLAHENRWDPDYFRTIGLALEDEDFIRIGHIVSPGTPCGNG 234
DB      169 GVMWMDSLASGTGLYINIMHD--WE--FEAMEIVLT-PDFL---PQVPEPTYLAG 218
OY      235 LTAQAAEKMILPGRTPVAVGLIDAHAGIGTGVEGGALNNLAVFGTSSCTMAST--- 290
DB      219 VKREYAEALMCIPEMLPFIIGSD---GALANIGIQAGQNDVITVGTSGAVKRLTDQF 275
OY      291 ---TSPFVPGWGPYY--SAMYPGMLVLEGGSAAGAALD-QLDFHRAVERARMAQ 343
DB      276 IDSRGRTFCYGAADGYFIAGAVNNGKVVEMGIDQFGSSEISRRDFASFIATIEV-- 333
OY      344 RVNQPPLPWLADRILEKTAOPSDAVALAKGLHVPEFLGNRAPDPHARAVICGLMER 403
DB      334 -----PPG-----AAGLLPQPYLIGERAPFWTDINGFVGL----- 365
OY      404 DLDNLALYI-AGLCGIGYGLRQILDAQTAQGVASKNIYVSGAGOHPLVRQIADTCGI 462
DB      366 TINHTKAHFIKRAILEGAVAFNLAEYEAVSAPDI---IYVTGGISADAMCKLADILNR 422
OY      463 PVITTOCCPVLIGSALGAVAGNIAPSVGEAMQOPTHDKYKYPQER-----YQ 512
DB      423 EIRVPHITIESSIGALAIKGRSLGILKDLN--LKHPLIKAVVHPSENLKVAELRLIFK 480
OY      513 SLNHRYEAVKQLOHTAK 530

```

Tue Mar 25 12:01:56 2003

us-09-802-208b-4.rag

Page 12

Db : : | | |
481 QVTTOLMSSYSQNLNSWOK 498

Search completed: March 13, 2003, 16:53:11
Job time : 46.7632 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 14.5877 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208b-4

Perfect score: 2811
Sequence: 1 MTITKTVIGVDVSGSVBAG.....HHRREAYKQLOHTAKLLRD 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_AA:*

1: /cgn2_6/protodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/protodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/protodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/protodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/protodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/protodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448.5	16.0	560	US-08-926-842B-13	Sequence 13, Appl
2	393.5	14.0	566	US-08-926-842B-18	Sequence 18, Appl
3	368.5	13.1	569	US-08-926-842B-19	Sequence 19, Appl
4	340	12.1	518	US-09-134-001C-4069	Sequence 4069, Ap
5	293.5	10.4	497	US-09-134-001C-4411	Sequence 4411, Ap
6	290.5	10.3	498	US-09-172-952-26	Sequence 26, Appl
7	263.5	9.2	501	US-09-172-952-15	Sequence 15, Appl
8	259	9.2	514	US-09-134-001C-4347	Sequence 4347, Ap
9	255.5	9.1	504	US-09-134-001C-4170	Sequence 4170, Ap
10	243.5	8.7	485	US-09-172-952-27	Sequence 27, Appl
11	167	5.9	709	US-08-968-563-18	Sequence 18, Appl
12	167	5.9	709	US-08-968-563-18	Sequence 18, Appl
13	167	5.9	709	US-08-968-563-18	Sequence 18, Appl
14	106.5	3.8	4928	US-09-370-700-5	Sequence 5, Appl
15	106.5	3.8	4928	US-09-370-700-5	Sequence 5, Appl
16	104.5	3.7	527	US-08-677-970-3	Sequence 3, Appl
17	103.5	3.7	731	US-08-911-364-1	Sequence 1, Appl
18	103.5	3.7	733	US-08-664-700-2	Sequence 2, Appl
19	102.5	3.6	527	US-08-907-229-2	Sequence 2, Appl
20	102.5	3.6	527	US-09-574-462-2	Sequence 2, Appl
21	102	3.6	792	US-08-678-039A-40	Sequence 40, Appl
22	101.5	3.6	480	US-09-108-020-6	Sequence 6, Appl
23	99	3.5	881	US-08-960-048-8	Sequence 8, Appl
24	98.5	3.5	1334	5476657-1	Patent No. 5476657
25	97.5	3.5	3739	US-09-320-878-2	Sequence 2, Appl
26	97.5	3.5	3739	US-09-105-537-33	Sequence 33, Appl
27	97.5	3.5	5087	US-09-144-085-1	Sequence 1, Appl

28	97.5	3.5	11877	4	US-09-105-537-6	Sequence 6, Appl
29	96	3.4	2475	4	US-09-413-814-48	Sequence 48, Appl
30	94.5	3.4	2152	4	US-09-036-987A-3	Sequence 3, Appl
31	94.5	3.4	2152	4	US-09-370-700-3	Sequence 3, Appl
32	93	3.3	612	2	US-08-746-283-31	Sequence 31, Appl
33	93	3.3	612	2	US-08-746-283-31	Sequence 31, Appl
34	93	3.3	613	2	US-08-746-283-1	Sequence 29, Appl
35	93	3.3	613	2	US-08-746-283-1	Sequence 1, Appl
36	93	3.3	613	2	US-08-746-283-5	Sequence 3, Appl
37	93	3.3	613	2	US-08-746-257A-1	Sequence 1, Appl
38	93	3.3	613	4	US-09-347-878-48	Sequence 48, Appl
39	93	3.3	700	4	US-09-408-647A-2	Sequence 2, Appl
40	93	3.3	1060	4	US-08-911-393-2	Sequence 2, Appl
41	92.5	3.3	1403	1	US-07-908-253-3	Sequence 3, Appl
42	92.5	3.3	1403	2	US-08-694-865-17	Sequence 17, Appl
43	92.5	3.3	1403	3	US-08-535-837-3	Sequence 3, Appl
44	92.5	3.3	1403	3	US-09-124-491-17	Sequence 17, Appl
45	92.5	3.3	2647	2	US-08-583-562B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-926-842B-13
Sequence 13, Application US/08926842B
Patent No. 6030807
GENERAL INFORMATION:
APPLICANT: Sa-No. 6030807ueira, Isabel
APPLICANT: de Lencastre, Herminia
TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentio Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926,842B
FILING DATE: 10-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-089 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
ORIGINAL SOURCE: protein
ORGANISM: Bacillus subtilis
FEATURE:
OTHER INFORMATION: /product= "arab"
US-08-926-842B-13
Query Match 16.0%; Score 448.5; DB 3; Length 560;
Best Local Similarity 27.1%; Pred. No. 3.2e-37;
Matches 153; Conservative 95; Mismatches 237; Indels 79; Gaps 20;

STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926,842B
FILING DATE: 10-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-089 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: *Salmonella typhimurium*
US-08-926-842B-19

Query Match 13.1%; Score 368.5; DB 3; Length 569;
Best Local Similarity 26.8%; Pred. No. 5.6e-29;
Matches 136; Conservative 79; Mismatches 235; Indels 113; Gaps 27;

8 IGVVSGSVAGIFDLSLSLHAT-EKITTTRSGSRVEQSS-----50
5 IGLDFGSDSVRALAVDC-----ATGDELATSVEMPRMNOEGYCCGRNNQFRHNRDXY 57
QY 51 QETWQAVSCIRNALTLADYCAOSVAGIGFDATCSL-VLDKKGDPRLVSPREGDAKON--107
DB 58 MESMEALKAUVALQLSAAQ--RANVVGIGVDSGTSPAPRIDAGDNVALRPEFENFNAM 115
QY 108 IIVMMHBRATEOARIATNH-----PVLNVGKISPEMETKILMLKEMPEIYRAG 162
DB 116 FVLMKDHAVEADEITRLCHKPKGYDSKRTIGIYSEFMWALHTVTRQDSAVQAAY 175
QY 163 OFEDLADFLTWATG-----DLARSYCTVTKWTMLAHENRMDPDYFRTIGL--AELADE 215
DB 176 SWIEICDMVALLSGTTRPRODIRGRSAGHKITLM--HES-WG-----GLRPSAFDE 225
QY 216 DFRIGHNIYSP-----GTPCGNGLTAQAAEMKGLRGTPTAVGLIDAHAGIGTVG 267
DB 226 LDPCINHLRYPLFSETFTADLPVGT-LCAEMAQRLDLPESVVISGAFCHMGAVGA--282
QY 268 VEGGALNNLALYVFGTSCCTMASTSPSF-----VPGVWGPRYYSAMVPRLMVEGOSAGA 323
DB 283 --GAQPTLVKVIQSTCDILADKOSVGRAYKIGIGOVDSVVPFRTIGLENGOSAFGD 340
QY 324 -----AIDQLDFRPAVEAREMAQRYNOPLRVMLADRLIEKTAQPSDAVLAK 372
DB 341 IVAMFSVILSWPLEQALQAPRELKQINASQK--QLLPAL-----IDAMAKNP 386
QY 373 GLHNVPE-----FLGNRPAPDPHARAVYGLGEMERDIDNLLALYIAGLCIGGLROILD 428
DB 387 SLPHLLPVLDLWFMGRTRPNRNORLKGVIITDNLNLTADAP--ALFGGLVASTAFARAIOE 443
QY 429 AQTAGGVSKNIYISGG-AGONPLVROILADTGIR--VITTOCCSEPVLLGSLGAVA 484
DB 444 CFTDQGIANNVNNMALGIAIKKNOYIMOVCCDVLRPLQIYASDGC--ALGAAIFRAVA 500
QY 485 GNIPASVGAEMOOF-THVDKYYYPQ-----ERYSLHRRREAY 522

Db 501 AKYHADIPAAQSGMASAVERTLRHPHOAQRFEOL-YRRYQOW 542

RESULT 4
US-09-134-001C-4069
Sequence 4069, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4069
LENGTH: 518
TYPE: PRT
ORGANISM: *Staphylococcus epidermidis*
FEATURE:
NAME/KEY: UNSURE
LOCATION: (6)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-134-001C-4069

Query Match 12.1%; Score 340; DB 4; Length 518;
Best Local Similarity 25.4%; Pred. No. 4.1e-26;
Matches 137; Conservative 85; Mismatches 236; Indels 82; Gaps 19;

25 NSGSLSHATEKITTRRSGRVDSOQETWQAVSCIRNALTLADYCAOSVAGIGFDATC 84
18 NGTYMNSLYDK-----PLGNYVFLQNMADDYLIQLEQGVFLDSKVNKNVVGIGVDFTS 73
QY 85 SLVY-LDKNDPLRVSPREGDAKON--IIVMMHBRATEOARI--NATHNPLVNVYSGK 137
DB 74 STIIFLDEQEPRL--NRHEDLKTPTPAVYVKLMKHNGADENYVIONSKKNWLYDYGSS 131
QY 138 ISPEMETKILMLKENPPEIYERAGOFEDLADFLTWATGDLARSYCTVTKWTMLAHEN 197
DB 132 VNSEMMIPIKLEKHNERRRARRYIMEAGDYITSLTNSNRNSGIGFKGW--DNEA 190
QY 198 RMDPYFRRTIGLELADEDFIRGINH-----IVSPGTCGNGLTAQAAEMKGLRGTPTVA 252
DB 191 GFNYDFEHSV-----DPDLPKIVKEKCEAPRTISIGESAGR-LCKDYQQTWGLSDYQVS 243
QY 253 VGLIDAHAGIGIVYVGGALNNLALYVFGTSCCTMASTSPSPVPGVWGPRYYSAMVPGLM 312
DB 244 PFIIDAHSGVAGVAGALEAG--EFTAVIGSTCHLMDSKQVPSISITGSKVKNAILIPGLY 300
QY 313 LVESGQSAAG-----AAIDQLDFH--PAVEAREMAQRYNOPLRVMLADRLIE 359
DB 301 AVEGQPAVVDLREYSKNQAPKHIVDOANHHMVLNLYLELSHI-----RIEE 350
QY 360 KTAQPSDAVALAKGLHNVPEPL--GNRAPDPHARAVYGLGEMERDIDNLLALYIAGLCG 418
DB 351 QHV-----VLLMLNGBRSILSNSHITGSIIFGLITQTPYEMIHRAVIE--A 394
QY 419 IGVGLROILDAQAGVSKNIYISGAGQ--HPLVROILADTGIRPITTOCCSEPVLLGS 477
DB 395 TAFETKLIMQFEDNHNRPVHTVYASGIPQSKLVEITVANNLKRKRVVYIDSSNAGSLGA 454
QY 478 AIIAGVANGNIAPSVGEA--MOQFTHVDKYUYPOERYOSLHRRREAYKOLQHTAKLRLD 534
DB 455 AMGANGNMYSTLKEALSMKQ-----PLAYIQEPIQVQAVKPLYNHNGYGLHD 505
RESULT 5
US-09-134-001C-4411

```
; Sequence 4411, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4411
; LENGTH: 497
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4411

Query Match      10.4%; Score 293.5; DB 4; Length 497;
Best Local Similarity 22.8%; Pred. No. 2.3e-21;
Matches 128; Conservative 83; Mismatches 249; Indels 101; Gaps 18;

QY 3 ITKTVIGDVGGSSVRAGIFDLNGLSLSHATEKITTRRSRVSQSOETMAVCSIR 62
DB 3 VKREVLDIDIGTSAIKTIADVQGNVIESVETLKYOEHPGSEODPNWEFEATKKGIK 62
QY 63 NALTLDVCAOSVAGIGFDATC-SLVVLDKNGDPLVPSPEGDAKONITVMDHATEQAE 121
DB 63 ELIOSTFMSKIVGIGSFGSMHGLVIYDNGIRPL-----RKAILMDTNSIQCR 113
QY 122 RINATHHPLVNVGKTSPEME---TPKILMKNEMPEIYERAGOFDLAFLTWRTGD 178
DB 114 QIEDIYERLNY-----NPLIEGFTLPKMLVQOHEPEIWRVDFMLPKDYLRCLTQT 168
QY 179 LARSCVITCKMTWLAHEN-RMPDVFRTIGLAELADEDFIRIGHIYSPCTPCGNGITA 237
DB 169 IHMEY-SDACSTLLEPNDEEMTRDVGDTFNIGDIYPP-----LVKSHSYGN-VTS 218
QY 238 QAAEMGLLPCTPVAVGLIDAHAGIG-----TWVEGALNNLAVYFGTSSTCT 286
DB 219 SLAKELDSSDVAVYAGGDNACGALIGAGYIHDKSALCSITSCVVL-NVEYQVNTISDS 277
QY 287 MASTSPSEVPGWGPRYSAMVPGMLVEGQSAAGALDQLDFHRAVEAREMAORVN 346
DB 278 NLH-----LFNHSVPTDYAMGYTLAAGYSLN----- 304
QY 347 QPLVWLADRILEKTA-----QPSDAVALAKGLHVPPELGNRAPFADPHARAVICGL 399
DB 305 -----WIKOTFFENESFEELINLAASSKIG-ANGLLFTPYLAGERTPHGDAQIIGSFIGI 358
QY 400 -GMEERDLNLALYIAGLCIGYGLRQILDQAOTAGVAVSKNIVISGAGOHPLRQIAD 458
DB 359 SCGHTKADFPARAV-----IEGITYLSYDSIKIMRAAGHEMNSITISGGAKSRFLQAD 414
QY 459 TCGIPVITTOCCPEVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYYPQ-----ERY 511
DB 415 IFNVOIKRLKHEEGPSMGAALAAVYGLGMFTKTESVCEAFIKYDEVPEPNENHDLVEOY 474
QY 512 QSLHHRVEAYKQLOHTAKLL 532
DB 475 YSYVEALYKQTKOL--TADLL 493

RESULT 6
US-09-172-952-26
; Sequence 26, Application US/09172952
; Patent No. 6368793
; GENERAL INFORMATION:
; APPLICANT: Hoch, James
; APPLICANT: Dattois, Veronique
; TITLE OF INVENTION: METABOLIC SELECTION METHODS
```

```
; FILE REFERENCE: 234/191
; CURRENT APPLICATION NUMBER: US/09/172,952
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 498
; TYPE: PR
; ORGANISM: LysK-Ec
US-09-172-952-26

Query Match      10.3%; Score 290.5; DB 4; Length 498;
Best Local Similarity 24.8%; Pred. No. 4.7e-21;
Matches 136; Conservative 77; Mismatches 243; Indels 93; Gaps 22;

QY 3 ITKTVIGDVGGSSVRAGIFDLNGLSLSHATEKITTRRSRVSQSOETMAVCSIR 62
DB 1 MTQVWLIGDGGSWLKKGLYDREGREGVQRLPICALSPQGMAREMDALMOCOMAVIR 60
QY 63 NALTLDVCAOSVAGIGFDAT-CSLVLDKNGDPLVPSPEGDAKONITVMDHATE--- 118
DB 61 ALLTHSGVSGBOIVGIGISAQKGLFLDKNDKPL-----GNAILSSDRAMEIYR 111
QY 119 --QAEKINATHNPVLN---YVGKISPEMETPKILMKNEMPEIYERAGOFDLAFLTW 173
DB 112 RMQEDGIREKLYPLRTROTLMGHPVS-----LIRWLKHEPERYAOIGCYMHTHYLRW 165
QY 174 RATGDLARSVTVCKMTWLAHEN-----RMPDVFRTIGLAELADEDFIRIGH--H 223
DB 166 CLTG-----VKCEESNISESNLYMNSLGEYRPLCDLWDIGLAE-----INHALRP 210
QY 224 IVSPGTCGNGLTQAAEMGLLPCTPVAVGLIDAHAGIGTVEEGALNNLAVYFGTS 283
DB 211 VVGSALICGE-ITQTAALGLKAGTAVVGGLPFVVSFAL-CAGIEDEFTLNA--VNGTW 266
QY 284 SCTMAST---TSPSEVPGWGPRYSAMVPGMLV-EGQSAAGALDQLDFHRAVEAR 339
DB 267 AVTSGITRGLRDEGAHNPVIGRYN---DGEFIYHESPTSSGN-----LEMPTA-QMG 317
QY 340 EMAORVNO---PLPVWLADRILEKTAQPSDAVALAKGLHVPPELGNRAPFADPHARAVI 396
DB 318 ISPEINQAVASLPKAGGDLFFLPFLGSMN-----GLEMTSGGYQMA-----INTRAHL 368
QY 397 CGLMEERDLNLALYIAGLCIGYGLRQILDQAOTAGVAVSKNIVISGAGOHPLRQI 456
DB 369 -----LQAIYE---GVVEFSHMTHLNRMRRERTDVHTLRVYGGRAHSDVMQM 413
QY 457 ADTCGIPVITTOCCPEVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYYPQOERYOSLHN 516
DB 414 ADVSGLRIELRPOVEETGCFGALARVGTGVYHNFSEBQRDLRHRPVRTLLPDMTAHOLYQ 473
QY 517 RRYEAYKOL 525
DB 474 KKYORYOHL 482

RESULT 7
US-09-172-952-15
; Sequence 15, Application US/09172952
; Patent No. 6368793
; GENERAL INFORMATION:
; APPLICANT: Hoch, James
; APPLICANT: Dattois, Veronique
; TITLE OF INVENTION: METABOLIC SELECTION METHODS
; FILE REFERENCE: 234/191
; CURRENT APPLICATION NUMBER: US/09/172,952
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 501
; TYPE: PR
; ORGANISM: LysK-Ko
```


US-09-172-952-15

Query Match 9.4%; Score 265.5; DB 4; Length 501;

Best Local Similarity 22.7%; Pred. No. 1.8e-18;

Matches 129; Conservative 74; Mismatches 224; Indels 141; Gaps 23;

```
OY 8 IGVNCGSVGAGIFDLNGLSLSHATEKITTTTRSGRSRVEOSSOEIOWAVCSIRNALTL 67
DB 9 LGIDCGGTLKAGLYBAEGHEGIVQALRTSPRLPGYERDMRQMLMOCAATITAGLLOQ 68
OY 68 ADVCAOSVAGIGFDT-CSLVLDKNGDPLVSPEDAKONITVMMDHRAE-----QAE 121
DB 69 AGVSGRQIKGVISAQGGFLDKODRPL-----GNALSSDRRLKIVGRMQRD 119
OY 122 RINATHHPVLN---YVGKISPEMETPKILMLKENMPETIERAGOFFDLADFLTWATGD 178
DB 120 RLPERLYPYRTROTLLMTGHPAS-----LIRWVENEPORQAIGCVMMGHDYLRWCLTG- 172
OY 179 LARSVCTVCKMTWLAHEN-----RMPDYERTIGLAELADEDEFIRIGHIIVSPGTPC 231
DB 173 -----AKGCEESNISSESNLYNMAMGOYDRLTEWLGIGETI-DSAL-----PPVYGSATC 221
OY 232 GNGLTAAOAAEMGLPRTPVAVGLIDAHAGIGTGVGEG-GALNMLAYVFGTSSCTMAST 290
DB 222 GE-ITAAALTLGLAGTPEVVGGLFDVYSTAL-CAGIEDESTLNAAVGTAVTSCIAHGL 279
OY 291 TSPSFYGVWGPYYSAMVFGMLVEGGQSAAGAIIDQLDFHRAVEAREMARVNQPLP 350
DB 280 RDHEAHPYVYGRYV--DQYIVHEASPTSSGNLE----- 312
OY 351 VWLADRIEKTQPSD-----AVA-LAKG---LHVPEFLGNRAFPADPHARAVICG 398
DB 313 -WF-----TAQMDLSPEIENQAVASLPKAGSELEFLPELIXGNSA-----G 352
OY 399 LGMERDLNLLALYIAGLCIGYGLROI-----LDAQTAGVSKNIV----- 441
DB 353 LEM-----TCGF-YGMOALHTRAHLLOAVEGVVFSHTHLRSRREFTNV 397
OY 442 -----ISSGAGOHPLVROILLADTCGIPVITTOCEPVLGSAIIGAVAGNIAPVGEAMQ 497
DB 398 QALRVITGGRAHSDVMOMLADVSGLRLELPKVEETGCGFALARVGTGYRSFSEARRA 457
OY 498 FTHVDRYVYQERYQSLHRRRYEAYKOL 525
DB 458 RQHPVRTLPLRDMTAHARQYRKRYHNYHL 485

RESULT 8
US-09-134-001C-4347
; Sequence 4347, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4347
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4347
```

Query Match 9.2%; Score 259; DB 4; Length 514;
Best Local Similarity 22.3%; Pred. No. 8.5e-18;
Matches 124; Conservative 82; Mismatches 232; Indels 118; Gaps 21;

```
OY 5 KTVIGDVGSGSVRAGIFDLNGLS-LTSH-----ATEKITTTTRSGRSRVEOSSOEIOWAV 57
DB 3 KVMIGVDGKTTSKAVLYDENCKFIKKNIGYDLHPNDVS-----FEENDELFDAY 55
OY 58 CSCIRNALTLADCAOSVAGIGFDTATC-SLVLDKNGDPLVSPEDAKONITVMMDHRA 116
DB 56 LMTVKYIVRESGIADKIDIFISLAQMHSILIANEDNORLTEN-----ITMDNRA 106
OY 117 TEOAERIMNTH--HPVLYNVGKISPEMETPKILMLKENMPETIERAGOFFDLADFLTMR 174
DB 107 NDVADLIERKSYGFEELYORTGPIHMSPLSKIFEMRHEBPKIFQTAPEADTKTVLQR 166
OY 175 -----ATGDLARSVCTVCKMTWLAHENRMPDYERTIGLAELADEDEFIRIG 221
DB 167 LYREFVIDYSLGATGMN-----LEOLDMOKALELIGIENQOLPOLY--- 210
OY 222 HHIVSPGTPGNGLTAAOAAEMGLPRTPVAVGLIDAHAGIGTGVGEGALNMLAYVFG 281
DB 211 -----PTTHVLTGMKKRYATLMGIDEQTPVIVGASD---GVLNLTGVSXQKGEVAVTIG 262
OY 282 TSSCTMASTSPSFYGVWGPYYSAMVFGMLVEGGQSAAGAIIDQLDFHRAVEAREM 341
DB 263 TSGAIFRTVINQPK--TDEKGRIFCYLLDKQYVIGGVNNGVVLN----- 306
OY 342 AQRVNQPLPWLADRIE--KTAQ-----PSDAVAL-----AKGLHVPEFLGNRA 385
DB 307 -----WLPDELLASVEETAKRLGVDPYVLQIASRVKPGABGLFHNYLGERA 356
OY 386 PRADPHARAVICGLGMRDLNLLALYIAGLCIGYGLROI-LDAQTAGVSKNIVISG 444
DB 357 PLNMADARGSFELTSLSHKKEHMIR--AALEGVLYNLVYVLLALIEVNMETPTTIKATG 413
OY 445 GAGOHPLVROILLADTCGIPVITTOCEPVLGSAIIGAVAGNIAPVGEAMQPTTHVD 502
DB 414 GFAKSETRWOMADIFDTDLVPESTYESSCLGACVLGAKALGEIDPSTIKDMVGTTHA- 472
OY 503 KYRYPQER---YQSL 514
DB 473 --HEPNEETVAITQOL 486
```

```
RESULT 9
US-09-134-001C-4170
; Sequence 4170, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4170
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4170
```

Query Match 9.1%; Score 255.5; DB 4; Length 504;
Best Local Similarity 21.0%; Pred. No. 1.9e-17;
Matches 117; Conservative 87; Mismatches 245; Indels 107; Gaps 20;

```
OY 5 KTVIGDVGSGSVRAGIFDLNGLSLSHATEKITTTTRSGRSRVEOSSOEIOWAVCSIRNA 64
DB 8 KYILSIDGTTSSRAILFNKEGELKVSQREKQHPHPWVEIDANEITVSLVSWAEL 67
OY 65 LTLADVCAOSVAGIGF-DATCSLVLDKN--GDPLVSPEDAKONITVMMDHRAEQAER 122
```

Db 68 LNENNINANOEGIGITNORETIVVWMDKNTGRPI-----YHAIYQSOROTODICTN 118
Qy 123 I-NATHHPLVNYGKISPEMPTKILMKENPEIYERAGO---FFDLADFLTWBAT 176
Db 119 LKEQGEYEFREKXTGLDLPYFAGTKYKMIIDHVGAREKANGDLLFGTIDSLWMLKLS 178
Qy 177 GDLARSVCTVCKMTWL--AHENRMDPYFRTIG-----LAELEDEDI---RICHNIYVS 226
Db 179 GRTHAITDYTNASRTLMENIYDLKWDDELLELNI PKOMLEPVKSSSEIYKTTIDYHFG 238
Qy 227 PGTPCGN-----GLTPQAAEMGLPGTPVAVGLIDAHNGIGCTVVEGGLNNLAVY 279
Db 239 QEVFIAGIAGDOQAALBEGQACFDKQDVKNYGTGCFMLMNGE--EAVASEGLTTITV- 296
Qy 280 FGTSSCTMASTSPSPVPGVWGPYYSANVPGMLVVEGQASAGAIDOLDLPHPAVEAR 339
Db 297 -----GLDGKVNVAL-----EGSIFVSGAIDQ----- 318
Qy 340 EMAQRVNOPLPVWLAD--RIEAKTAQP---SDAVALAKGLHVPEFLGNRAPFADPHARA 394
Db 319 -----WLRDGLRMINSAPQTEYNASRVSESTEGVYVMPAFVGLGTPYMDSEARG 366
Qy 395 VIGGLMERDNLNLAATYIAGLCIGYGLRQIIDAQTA--QGVVSKNIYISGAGOHPLVR 453
Db 367 AIFGLSKTEKEHPIRATLESCL---YOTRDVMEAMSKDSGIEVONLKVGDGAANNFTM 423
Qy 454 QILADTCGIPYITTOCCPEVLLGSAILGAVAGNIAPSVG-----EAMQOFTHVDKYYPQ 508
Db 424 QFOADIYVSSVEREIOETTLAAGAYLAGLA-----VGFMDKEDIERKKLOTEFRP- 476
Qy 509 ERYOSLHRRREAVKQ 524
Db 477 EMDADORHKLYSGWK 492

RESULT 10
US-09-172-952-27

Sequence 27, Application US/09172952
Patent No. 6368793
GENERAL INFORMATION:
APPLICANT: Hoch, James
TITLE OF INVENTION: METABOLIC SELECTION METHODS
FILE REFERENCE: 234/191
CURRENT APPLICATION NUMBER: US/09/172.952
CURRENT FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 485
TYPE: PRT
ORGANISM: LYXK-H1
US-09-172-952-27

Query Match 8.7%: Score 243.5; DB 4: Length 485;
Best Local Similarity 23.2%: Pred. NO. 3e-16;
Matches 129; Conservative 77; Mismatches 242; Indels 109; Gaps 22;

Qy 8 IGVVGVSGSVRAGIFDINGSLSLHATEKITTRRSQSVESQSOIOWAVCSITRNALTL 67
Db 5 LGIDCGGTFIKALIFDONGTQIARRNIPITSEKPGIAEADMDELINLCAQVIOKTRQ 64
Qy 68 ADVCAQSVAGIGFDT--CSLVYLDKNSDPLVSPEDAKONIYVMDHRAE-----QAE 121
Db 65 SSILPQOIKALIGISAQGAFFLDKNKPL-----GRA-----LSSDRAVEIYQCMQKE 115
Qy 122 RINTTHPV---LNYVGKISPEMPTKILMKENPEIYERACOFDLDPLTWRTGD 178
Db 116 NIIQKFPYITLQTLMMGPVS-----ILRWIKENEPSRYBOJITILMSHYLEFCLNEK 169
Qy 179 LARSVCYVTCWKMTLAHNNRMDPYFRTIGLAELADEFLIGHNIYSP---GTPCGML 235
Db 170 LYCEETNISSENFYNNMBGKIDQLAKLFGITTECDK-----LPIITKSNKIAGYV 220

Qy 236 TAQAAEMGLPGTPVAVGLIDAHAGIGTVGEGAL-----NLAIVYFGTSSCTMA 288
Db 221 TSRAEQSGIVEGIFPVVGLFD-----VSTALCADLKDQDOHLNVLGTWSVVG 270
Qy 289 ST---TSPSPVPGVWGPYYSANVPGMLVVEGQASAGAIDOLDLPHPAVEAREMQRV 345
Db 271 VTHYIDNQTIPFVYGKPKP-----NKFTIHESPTSGNLEWV 311
Qy 346 NO-PLPVMLADRI---LEKTAQPSDAVALAKGLHVPEFLGNRAPFADPHARAIVCGLM 401
Db 312 NOFNLPN--VDINHEIKLKPASSVLPAPFLYSNNKLGMOGF-----YGIQS 360
Qy 402 ERDDNLL-ALYIAGLCIGYGLRQIIDAQTAQGVVSKNIYISGAGOHPLVRQIADTC 460
Db 361 HHTQHLLOAIYE-----GVIFSMHSLEBRMQVRFPNASTVRTGSPAKSEVMQMLADIS 416
Qy 461 G-----IPYITTOCCPEVLLGSAILGAVAGNIAPSVGEMQOFTYVD-KYTPQERTQSLH 515
Db 417 GMRLEIPNIEETGC---LGAALMAQESAV-----BISQILNIDRKIFLPDKNQYSKY 467
Qy 516 HRYEAYKQLOHTAKL 532
Db 468 QKHRYIKFIEALKNL 484

RESULT 11
US-08-968-563-18

Sequence 18, Application US/08968563
Patent No. 6013494

GENERAL INFORMATION:
APPLICANT: CHARLES E. NAKAMURA
APPLICANT: ANTHONY A. GATENBY
APPLICANT: AMY (KUNAG-HUA) HSU
APPLICANT: RICHARD D. LA BEAU
APPLICANT: SHARON L. HAYNIE
APPLICANT: MARIA DIAZ-TORRES
APPLICANT: DONALD E. TRIMBUR
APPLICANT: GREGORY M. WHITED
APPLICANT: VASANTHA NAGARAJAN
APPLICANT: MARK S. PAYNE
APPLICANT: STEPHEN K. PICATAGGIO
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 4 CAMBRIDGE PLACE
STREET: 1870 SOUTH WINTON ROAD
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 INCH DISKETTE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968.563
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY

```

: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: CR-9982
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 302-892-8112
: TELEFAX: 302-773-0164
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 709 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: GUT1
: US-08-968-563-18

```

```

Query Match          5.9%; Score 167; DB 3; Length 709;
Best Local Similarity 21.0%; Pred. No. 4,2e-08;
Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25;

```

```

OY 46 VEOSSOEIMQAVSCIRNA-LTLADYCAQSV-----GIGFDATCSLVLDKNG 93
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 153 VEHPOKLLVNVYQCLASSLSLQITNSERVANGLPYKYICGIANMRETTILMSRTG 212
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 94 DPLVSPBGDAKONIIVMDHRAE-----QAEINATGHPVLYVGGKISP 140
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 213 KPI-----VYGIWMNDTRIKIVRDKWMTSVDRQLRQKTGLPRL-----ST 257
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 141 EMETPKILMKEMP---EYERAGOFEDLADFLTW-----RATGDLARSVCTVTC 188
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 258 YFSCSKIRWFLDNEPLCTKAYEENDLMTFTVD--TWLIYQLTOKAFVSQVTNASRTGFM 315
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 189 KWTWLAHENR---WDPD-----YFTIGLAELADEDFIRIGH-----223
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 316 NLSTLKYDNELLEFWGIDKRLIMPELVSSQYRGDCI-----PDMIMKLDSPKTVL 370
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 224 --IVSPGTCGNGLTAQAAEMGLPGTPVA-----VGLIDHAGSIGTVYEGGALLN 275
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 371 RDLVKKRNLPIQCGIDQASAMVGLAYKPGAACCTVGTGCFLLYNGTKKLISOHGALT 430
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 276 LAVVFTSSCTMASTTSPSPVPGWGPY---YSAMVPGCL---WLVGGGSAAGAALDQL 328
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 431 LAF-----WEPHLOEYGGQKPELSKPHFALEGSVAVAGAVVO--467
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 329 LDFHPAVEAREMAQRYNQPLPYWLAD--RILEKTAQ---SDAVALAKGHLVHVPFELGN 383
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 468 -----WLRDNLRLIDKSEVGVRIASTVDPGSGVVPVPAFSGL 504
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 384 RAFPADPHARAIVICGLAMERDLDNLALYIAGLCIGYGLRQLIDAQT--AQGVSKN--439
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 505 FAYWMPDARATIMGSOFTTASHIARAAGEVC---FQARAILIKMSSDAFEGSGKDRD 561
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 440 -----IYISGAGQHPLVRLQDLADTCGIPVY-----TTQCCPEVLIG 476
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 562 FLEISDVYKESPLSVLAVDGMGMSRNEVMOIADILG-PCVKVRRSPTAEC-----613
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 477 SAILGAVAGNIA 488
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 614 TALGAALIAANMA 625
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 12
US-08-969-683A-18
: Sequence 18, Application US/08969683A
: Patent No. 6136576
: GENERAL INFORMATION:
: APPLICANT: GENENCOR INTERNATIONAL, INC.
: TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genencor International, Inc.
: STREET: 4 Cambridge Place

```

```

: STREET: 1870 South Winton road
: CITY: Rochester
: STATE: NY
: COUNTRY: U.S.A
: ZIP: 14618
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSeq for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/969,683A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US97/20873
: FILING DATE: 13-NOV-1997
: APPLICATION NUMBER: 60/030,601
: FILING DATE: 13-NOV-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Glaister, Debra
: REGISTRATION NUMBER: 33,888
: REFERENCE/DOCKET NUMBER: GC 369-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-864-7620
: TELEFAX: 650-845-6504
: TELEX:
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 709 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: GUT1
: US-08-969-683A-18

```

```

Query Match          5.9%; Score 167; DB 4; Length 709;
Best Local Similarity 21.0%; Pred. No. 4,2e-08;
Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25;

```

```

OY 46 VEOSSOEIMQAVSCIRNA-LTLADYCAQSV-----GIGFDATCSLVLDKNG 93
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 153 VEHPOKLLVNVYQCLASSLSLQITNSERVANGLPYKYICGIANMRETTILMSRTG 212
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 94 DPLVSPBGDAKONIIVMDHRAE-----QAEINATGHPVLYVGGKISP 140
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 213 KPI-----VYGIWMNDTRIKIVRDKWMTSVDRQLRQKTGLPRL-----ST 257
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 141 EMETPKILMKEMP---EYERAGOFEDLADFLTW-----RATGDLARSVCTVTC 188
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 258 YFSCSKIRWFLDNEPLCTKAYEENDLMTFTVD--TWLIYQLTOKAFVSQVTNASRTGFM 315
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 189 KWTWLAHENR---WDPD-----YFTIGLAELADEDFIRIGH-----223
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 316 NLSTLKYDNELLEFWGIDKRLIMPELVSSQYRGDCI-----PDMIMKLDSPKTVL 370
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 224 --IVSPGTCGNGLTAQAAEMGLPGTPVA-----VGLIDHAGSIGTVYEGGALLN 275
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 371 RDLVKKRNLPIQCGIDQASAMVGLAYKPGAACCTVGTGCFLLYNGTKKLISOHGALT 430
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 276 LAVVFTSSCTMASTTSPSPVPGWGPY---YSAMVPGCL---WLVGGGSAAGAALDQL 328
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 431 LAF-----WEPHLOEYGGQKPELSKPHFALEGSVAVAGAVVO--467
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 329 LDFHPAVEAREMAQRYNQPLPYWLAD--RILEKTAQ---SDAVALAKGHLVHVPFELGN 383
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 468 -----WLRDNLRLIDKSEVGVRIASTVDPGSGVVPVPAFSGL 504
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 384 RAFPADPHARAIVICGLAMERDLDNLALYIAGLCIGYGLRQLIDAQT--AQGVSKN--439
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 505 FAYWMPDARATIMGSOFTTASHIARAAGEVC---FQARAILIKMSSDAFEGSGKDRD 561
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

OY	440	-----IYISGACOHPLVRQILADTCIPVI-----TTQSCBPVLG	476
DB	562	FLEELSDTYEKSPSLVLAVDGMSRNSNEVMQIDILG-PCVKRRSTAE	613
OY	477	SAILCAVAGNIA 488	
DB	614	TALGAATANMA 625	
		RESULT 13	
		US-09-297-928-15	
		; Sequence 15, Application US/09297928	
		; Patent No. 6358716	
		GENERAL INFORMATION:	
		APPLICANT: BULTHUIS, BEN A.	
		GATENBY, ANTHONY A.	
		HAYNIE, SHARON L.	
		HSU, AMY K.	
		LAREAU, RICHARD D.	
		TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF	
		GLYCEROL BY RECOMBINANT	
		ORGANISMS	
		NUMBER OF SEQUENCES: 25	
		CORRESPONDENCE ADDRESS:	
		ADDRESSEE: GENENCOR INTERNATIONAL, INC.	
		STREET: 4 CAMBRIDGE PLACE	
		1870 SOUTH WINTON ROAD	
		CITY: ROCHESTER	
		STATE: NEW YORK	
		COUNTRY: U.S.A.	
		ZIP: 14618	
		COMPUTER READABLE FORM:	
		MEDIUM TYPE: DISKETTE, 3.5 INCH	
		COMPUTER: IBM PC COMPATIBLE	
		OPERATING SYSTEM: MICROSOFT WINDOWS 95	
		SOFTWARE: MICROSOFT WORD VERSION 7.0A	
		CURRENT APPLICATION DATA:	
		APPLICATION NUMBER: US/09/297,928	
		FILING DATE: 11-MAY-1999	
		CLASSIFICATION: <unknown>	
		PRIOR APPLICATION DATA:	
		APPLICATION NUMBER: 60/03602	
		FILING DATE: NOVEMBER 13, 1996	
		ATTORNEY/AGENT INFORMATION:	
		NAME: FLOYD, LINDA AXAMETHY	
		REGISTRATION NUMBER: 33,692	
		REFERENCE/DOCKET NUMBER: CR-9981-P1	
		TELECOMMUNICATION INFORMATION:	
		TELEPHONE: 302-892-8112	
		TELEFAX: 302-773-0164	
		INFORMATION FOR SEQ ID NO: 15:	
		SEQUENCE CHARACTERISTICS:	
		LENGTH: 709 amino acids	
		TYPE: amino acid	
		STRANDEDNESS: unknown	
		TOPOLOGY: unknown	
		MOLECULE TYPE: protein	
		SEQUENCE DESCRIPTION: SEQ ID NO: 15:	
		US-09-297-928-15	
		Query Match	
		Best Local Similarity 5.9%; Score 167; DB 4; Length 709;	
		Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25	
OY	46	VEOSQEIMQAVCSIRNA-LTLADVCAOSVA-----GIGFDATCSLVLDKNG	93
DB	153	VECHQPKLLVVNVOCIASLSLQFINSERVANGLPKYVICMGIANMBRETIILMSRTG	212
OY	94	DPLRPVSPGDAAQNIIIVMNDHRATE-----QAERIINNTHNPVLNVYGCKISP	140
DB	213	KPI-----VNGCIVMDTFTITIVRDKMONTSVDRQLRLRKGTGLELL-----ST	257

[illegible]

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4928 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-036-987A-5

Query Match 3.8%; Score 106.5; DB 4; Length 4928;
 Best Local Similarity 23.8%; Pred. No. 1.7;
 Matches 126; Conservative 65; Mismatches 201; Indels 137; Gaps 31;

QY 34 EKITTTT-----GSRVEQS-----SQE--IWQAVSCIRNALTLADY 70
 Db 2031 ERLSDARNRHEVLAVYRGSAVNODGASNGLTAPNGPSQQRVTIQTALTS---GLSVSDV 2087
 QY 71 CAQSVAGIG-----FDATCSLYVDKNGDPLPVSPGDAKONIYVMDIRATQARIN 124
 Db 2088 DAVEAHGTGRLGDPLEAQAALATYGRDRPGRLPLMGVSKSNT----- 2131
 QY 125 ATNHPVLNYGKISPEM-----ETPKILMKEMPEIYRAGOFEDLADFLWRATGDL 179
 Db 2132 -GHTQAAGAGYIKKMMARQGLPRTLHYDEPSAOVDSAGTVOLLTENTPDPDSGRL 2190
 QY 180 ARS-VCTVTCKWTWLAH-----ENRMDPDYFRTIGLAELADEDFTRIGHIIVSPG 228
 Db 2191 RRAGVSSFGISGT-NAHLILEQPPRESQSRSTED-----SGSVRDF-PVYPMWVSGK 2240
 QY 229 TPCNGTLTAQAAEMGLLPCTPVAVGLIDAHAGIG-TVYEGGALNNLAVYFTSSCTM 287
 Db 2241 TP--EALSAQADALMSYLSNR-----VDASPRDIGYSLAVTRPALDHRRAVVLGADRAAL 2292
 QY 288 -----ASTSPSPVPGWGPYYSAMVPGLMLVEGGO-SAGGAALDQLLD-FHPAVEBA-- 338
 Db 2293 LPLKALAVSNDAAEVITGTRAAQPGVGFVSGGGQMPGMSGLHSAPVFAADFDEACC 2352
 QY 339 -----REMAQRVNOPLPVWLAD-----RILEKT--AOPSDAVALAKGLHVVPEFLGN 383
 Db 2353 ELDAHLCQMAR-----LRDVLSSGSDQLLDQTLMAQPG-LFALQVGLM--ELLGS 2399
 QY 384 RAPADPHARAIVICGLMERDNLNLAITYAGLCIGIGYGLRQIID-AQTAGQVSKNIVI 442
 Db 2400 WGV-----RPVVVLG-----HSVGETLAFAAGVLSLRDAARLVAGRRLMQALPTGGAML 2450
 QY 443 SGGAGQHPVROLADTCGIPVITTOCCP---VLGSA-ILGAVAGNI 487
 Db 2451 AAAAGEBQL-RPLAD-CGDRVGIATAVNAFGSVVLSGDRVDLDDIAGRLL 2497

RESULT 15

US-09-370-700-5
 ; Sequence 5, Application US/09370700
 ; Patent No. 6274350

; GENERAL INFORMATION:

; APPLICANT: Baltz, Richard H
 ; APPLICANT: Broughton, Mary C
 ; APPLICANT: Crawford, Kathryn P
 ; APPLICANT: Madduri, Krishnamurthy
 ; APPLICANT: Treadway, Patti J
 ; APPLICANT: Turner, Jan R
 ; APPLICANT: Waldron, Clive
 ; TITLE OF INVENTION: Biosynthetic Genes for Spinosyn Insecticide
 ; FILE REFERENCE: 50489 DIV1
 ; CURRENT APPLICATION NUMBER: US/09/370,700
 ; CURRENT FILING DATE: 1999-08-09
 ; EARLIER APPLICATION NUMBER: US 09/36987
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 4928

; TYPE: PRT
 ; ORGANISM: Saccharopolyspora spinosa

US-09-370-700-5

Query Match 3.8%; Score 106.5; DB 4; Length 4928;
 Best Local Similarity 23.8%; Pred. No. 1.7;
 Matches 126; Conservative 65; Mismatches 201; Indels 137; Gaps 31;

QY 34 EKITTTT-----GSRVEQS-----SQE--IWQAVSCIRNALTLADY 70
 Db 2031 ERLSDARNRHEVLAVYRGSAVNODGASNGLTAPNGPSQQRVTIQTALTS---GLSVSDV 2087
 QY 71 CAQSVAGIG-----FDATCSLYVDKNGDPLPVSPGDAKONIYVMDIRATQARIN 124
 Db 2088 DAVEAHGTGRLGDPLEAQAALATYGRDRPGRLPLMGVSKSNT----- 2131
 QY 125 ATNHPVLNYGKISPEM-----ETPKILMKEMPEIYRAGOFEDLADFLWRATGDL 179
 Db 2132 -GHTQAAGAGYIKKMMARQGLPRTLHYDEPSAOVDSAGTVOLLTENTPDPDSGRL 2190
 QY 180 ARS-VCTVTCKWTWLAH-----ENRMDPDYFRTIGLAELADEDFTRIGHIIVSPG 228
 Db 2191 RRAGVSSFGISGT-NAHLILEQPPRESQSRSTED-----SGSVRDF-PVYPMWVSGK 2240
 QY 229 TPCNGTLTAQAAEMGLLPCTPVAVGLIDAHAGIG-TVYEGGALNNLAVYFTSSCTM 287
 Db 2241 TP--EALSAQADALMSYLSNR-----VDASPRDIGYSLAVTRPALDHRRAVVLGADRAAL 2292
 QY 288 -----ASTSPSPVPGWGPYYSAMVPGLMLVEGGO-SAGGAALDQLLD-FHPAVEBA-- 338
 Db 2293 LPLKALAVSNDAAEVITGTRAAQPGVGFVSGGGQMPGMSGLHSAPVFAADFDEACC 2352
 QY 339 -----REMAQRVNOPLPVWLAD-----RILEKT--AOPSDAVALAKGLHVVPEFLGN 383
 Db 2353 ELDAHLCQMAR-----LRDVLSSGSDQLLDQTLMAQPG-LFALQVGLM--ELLGS 2399
 QY 384 RAPADPHARAIVICGLMERDNLNLAITYAGLCIGIGYGLRQIID-AQTAGQVSKNIVI 442
 Db 2400 WGV-----RPVVVLG-----HSVGETLAFAAGVLSLRDAARLVAGRRLMQALPTGGAML 2450
 QY 443 SGGAGQHPVROLADTCGIPVITTOCCP---VLGSA-ILGAVAGNI 487
 Db 2451 AAAAGEBQL-RPLAD-CGDRVGIATAVNAFGSVVLSGDRVDLDDIAGRLL 2497

Search completed: March 13, 2003, 16:42:17
 Job time : 21.5877 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 15.0298 Seconds

(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208b-4

Perfect score: 2811

Sequence: 1 MTTKTVIGVDVSGSVRAG.....HRRYEVAKQLOHTAKLRD 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2811	100.0	534	9	US-09-802-208b-4
2	306	10.9	484	10	US-09-815-242-10383
3	303	10.8	501	10	US-09-815-242-10866
4	302.5	10.8	501	10	US-09-815-242-11080
5	289	10.3	517	10	US-09-815-242-12837
6	288	10.2	509	10	US-09-815-242-5796
7	279.5	9.9	509	9	US-09-738-626-6686
8	276.5	9.8	502	10	US-09-815-242-10425
9	276.5	9.8	502	10	US-09-815-242-11431
10	273.5	9.7	502	10	US-09-815-242-13660
11	267.5	9.5	505	9	US-10-116-821-14
12	264.5	9.4	496	10	US-09-815-242-5321
13	264.5	9.4	498	10	US-09-815-242-12335
14	262.5	9.3	501	10	US-09-815-242-13788
15	254	9.0	505	10	US-09-815-242-11927
16	246.5	8.8	496	10	US-09-815-242-10714
17	221	7.9	473	9	US-09-738-626-3630
18	219.5	7.8	447	10	US-09-815-242-11747
19	187.5	6.7	494	9	US-09-738-626-6829

20	167	5.9	709	9	US-09-308-207-18	Sequence 18, Appl
21	164	5.8	518	12	US-10-001-852-12	Sequence 12, Appl
22	108.5	3.9	513	10	US-09-833-745-63	Sequence 63, Appl
23	107	3.8	439	9	US-09-820-843A-60	Sequence 60, Appl
24	104.5	3.7	527	9	US-10-176-640-3	Sequence 3, Appl1
25	101.5	3.6	63	10	US-09-764-847-580	Sequence 580, App
26	101.5	3.6	650	10	US-09-815-242-13341	Sequence 13341, A
27	101.5	3.6	2993	9	US-09-738-626-6239	Sequence 6239, Ap
28	101	3.6	943	10	US-09-815-242-12027	Sequence 12027, A
29	100.5	3.6	438	9	US-09-738-626-4295	Sequence 4295, Ap
30	100.5	3.6	729	10	US-09-287-849-2	Sequence 2, Appl1
31	99	3.5	881	10	US-09-838-539-8	Sequence 8, Appl1
32	97.5	3.5	570	9	US-09-738-626-5603	Sequence 5603, Ap
33	97.5	3.5	3739	9	US-09-860-846-33	Sequence 33, Appl
34	97.5	3.5	3739	10	US-09-861-289-33	Sequence 33, Appl
35	97.5	3.5	11877	9	US-09-860-846-6	Sequence 6, Appl1
36	97.5	3.5	11877	10	US-09-861-289-6	Sequence 6, Appl1
37	97	3.5	701	10	US-09-900-237-32	Sequence 32, Appl
38	96.5	3.4	403	10	US-09-996-194-6	Sequence 6, Appl1
39	96.5	3.4	640	10	US-09-815-242-10786	Sequence 10786, A
40	95.5	3.4	1827	9	US-09-712-363-261	Sequence 261, App
41	95.5	3.4	4999	9	US-09-976-059-15	Sequence 15, Appl
42	93.5	3.3	506	10	US-09-900-237-20	Sequence 20, Appl
43	93	3.3	1060	10	US-09-955-909-2	Sequence 2, Appl1
44	92.5	3.3	740	9	US-10-051-909-37	Sequence 37, Appl
45	92.5	3.3	1080	10	US-09-900-237-30	Sequence 30, Appl

ALIGNMENTS

```
RESULT 1
US-09-802-208b-4
; Sequence 4, Application US/09802208B
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parrott, Wayne
; APPLICANT: Lafayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Arabidol or Ribitol As Positive Selectable Markers
; FILE REFERENCE: UGA-855R
; CURRENT APPLICATION NUMBER: US/09/802,208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-802-208b-4

Query Match      100.0%; Score 2811; DB 9; Length 534;
Best Local Similarity 100.0%; Pred. No. 7.1e-251;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTTKTVIGVDVSGSVRAGIFDNGSLSHATEKITTRSSGRVROSOEIMQAVCSC 60
    |||||||
DB 1 MTTKTVIGVDVSGSVRAGIFDNGSLSHATEKITTRSSGRVROSOEIMQAVCSC 60
    |||||||

OY 61 IRNALTLADYCAQSVAGIGDATCSLVLDKNDPLFVSPGDAKONIYMDHRAEQ 120
    |||||||
DB 61 IRNALTLADYCAQSVAGIGDATCSLVLDKNDPLFVSPGDAKONIYMDHRAEQ 120
    |||||||

OY 121 ERINATHHPVLNLYVGGISPEMETPKILMKENNPETLERAQGFEDLADFLTRATCDLA 180
    |||||||
DB 121 ERINATHHPVLNLYVGGISPEMETPKILMKENNPETLERAQGFEDLADFLTRATCDLA 180
    |||||||

OY 181 RSVCTVYCKMTWLAEHRMPDPYFRTIGLAELDEDEIRIGHIIVSGTCGNGLTROAA 240
    |||||||
DB 181 RSVCTVYCKMTWLAEHRMPDPYFRTIGLAELDEDEIRIGHIIVSGTCGNGLTROAA 240
    |||||||

OY 241 AEMGLLPGRVPAVAGLIDAHAGIGTGVEGGALNNLAIVFGTSCTMASTTSFVPGVW 300
    |||||||
```

```
Db 241 AEMGLGCTPAVGLIDAHAGGCTVGVGGALNNLAIVETSSCTMASTTSPSFGVGV 300
Qy 301 GPPYSAMVPGIMLVGGOSAGAAIDLDLPHPAVEAREMAQVNOPLPWLADRILEK 360
Db 301 GPPYSAMVPGIMLVGGOSAGAAIDLDLPHPAVEAREMAQVNOPLPWLADRILEK 360
Qy 361 TAQSSDAVALAKGLHVPPEFLGNRPADPHAAVIGCLGEMERLDNLALYINGLCIG 420
Db 361 TAQSSDAVALAKGLHVPPEFLGNRPADPHAAVIGCLGEMERLDNLALYINGLCIG 420
Qy 421 YGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTGTPIYITTOCCPEVLLGSAIL 480
Db 421 YGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTGTGTPIYITTOCCPEVLLGSAIL 480
Qy 481 GAVAGNIAPSVGEAMQOFTHVDKYYPQERYQSLHHRREAYKOLQHTAKLLRD 534
Db 481 GAVAGNIAPSVGEAMQOFTHVDKYYPQERYQSLHHRREAYKOLQHTAKLLRD 534

RESULT 2
US-09-815-242-10383
; Sequence 10383, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10383
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10383

Query Match 10.9%; Score 306; DB 10; Length 484;
Best Local Similarity 24.4%; Pred. No. 6.1e-20;
Matches 132; Conservative 75; Mismatches 249; Indels 86; Gaps 17;
```

```
Db 106 TLLEARVPOSRTVTGNLMPEGFTPAKLIWQNRHEPEIFRQIDKVLKROVILRMGTGEFA 165
Qy 181 RSVCTVTCCKMTWLAHENR-WDPDFERTTGLAELADEDFIRIGHIIVSPGPGNGLTAAQ 239
Db 166 -SDMSDAAGTMTLVNARQDSVWLQAC-----DLRQMPALVEGSEITGALLPEV 216
Qy 240 AAENGLLPGTFVAVGLIDAHAGGIGTVGVGGALNNLAIVFGTSSCTMASTTSPSPVGV 299
Db 217 AKANGMAT-PPVVVGGGDNMAGAVGVMDA---NQAMLSLGTGVYFA--VSEGLFSKP 270
Qy 300 WGPYYS--AMVPGIMLVGGOSAGAAIDOL-----LDHPAVEAREMAQVNOPLPWL 352
Db 271 ESAVHSCHAIPQRNHLMSVWLSASCLDMAKLTGSLNVPALAAQAQADESHE--PVM 328
Qy 353 LADRILEKTAQPSDAVALAKGLHVPPEFLGNRPADPHAAVIGCLGEMERLDNLALY 412
Db 329 F-----LPTLSGERTRHNHPQAKGVFEGLTHQGPHELAR-- 363
Qy 413 IAGLCGIGYGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTGTG--IPYITTOCC 470
Db 364 -AVLEGVYALADMDVNHACGIRPQSVTLIGGARSEYWRQMLADISGOOLDYRTGADV 422
Qy 471 EPVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYPQERYQSLHHRREAYKOLQHTAK 530
Db 423 GPA-LGAARLAQIDANPEKSLIELLPOL-----PLEQSHLPDAQRYAAYOPRRETFR 473
Qy 531 LL 532
Db 474 RL 475
```

```
RESULT 3
US-09-815-242-10866
; Sequence 10866, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10866
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10866

Query Match 10.8%; Score 303; DB 10; Length 501;
Best Local Similarity 23.4%; Pred. No. 1.2e-19;
```



```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11080
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11080

Query Match      10.8%; Score 302.5; DB 10; Length 503;
Best Local Similarity 23.3%; Pred. No. 1,4e-19;
Matches 128; Conservative 76; Mismatches 255; Indels 87; Gaps 18;

OY      1 MTTTTCVIGVDVSSVSRAGIFDLNGLSLSHAATEKITTTTNRSGSRVEDSSQEIQAVCSG 60
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      1 MTKDKYIILADGGTSSRAVLDDHNANVEIAQREFTQIYPRAGWEHNPMEIMATQSST 60
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      61 IRNALITLADVCAQSASVAGIGFDATGCLVYLDRKNGDPLPSPBEDAQNITVYMDHRTAEBA 120
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      61 LNEVAKAGITSDIEIAIGITNQRETTTWEKSTGTPV-----YNALVWCRRRTADTT 113
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      121 EIRINTNHP--VLNVYVGKISPEMETPKILMKENNP---EIERAGQFPLDAD--FLTWK 174
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      114 DKLKADGHEEIRNTTGLVNDYFESGTAKYKWLVDNEGAREKAEKAGELLFGVDTWLVWK 173
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      175 ATGDARSVCVTCKMTWL--AHENRMDPDYFRTIGLAELADEDFIRIGHIVSPGTCG 232
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      174 LTQGVHVTIDTYNASTRTMLFNHTKQMDKMLEILINIPRSMLE--VRNSETIYGQTNIGG 232
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      233 NGLTQAAAAEMKLLFGTPAVGLI--DAHAGSIGVYVGAGLNNLNAVFGSSCTMAST 290
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      233 KG-----GVRIPVAGIAGDQDAALYGLCVHAGGAKN--TYGT-GCFMLIH 275
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      291 TSPSPVPGVWG-----PYRSANVPGLMLVEGSGSAAGAIIDQLDFHPAVEEA 338
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      276 TGNKAKITSNGLLTTIACAKAPEPA-----LEGSVFTAGASIQMLRBLKTVHDS 327
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      339 RE--MAQRVNQPLPWLADRILEKTAQPSDAVALAKGLHVVPDFLGNRAPFADPHARAV 395
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      328 FDSFEYFAQVTD-----SNGYVVPAPFGLGAPWMDPYARBA 364
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      396 ICGLGERDLDLALYIAGLGGIGYGLRQIILDA--QTAGVYSKNIVISGGAGQHPVYQ 454
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      365 IFGLSRGANRNHIVR--ATLESIAVYQTDVYLEAMQSDGSERLQYIRYVGGATNNFNLQ 421
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      455 ILADPCGIVITTOCEPPLIGSAILGAAVAGNIAPSVGAMQOFTHVDYXYPOERYOSL 514
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      422 FQADLLDVAVERPVEYKAYALGAATLGIATGFWKDDLE--LDRKARVEXTFSPDSNEK- 479
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY      515 HHRRYEAYKQ 524
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      480 RRRRYKGMWK 489
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 5
US-09-815-242-12837
; Sequence 12837, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12837
; LENGTH: 517
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12837
```

Query Match 10.3%; Score 289; DB 10; Length 517;

Best Local Similarity 22.9%; Pred. No. 2, 5e-18; Matches 131; Conservative 92; Mismatches 244; Indels 104; Gaps 23;

```

QY 5 KTVIGVDSGVSRAIGFDLNGS-LLSHA-----TEKITTRSGSRVSSOEIMQAV 57
DB 2 KYMIGVIGITSTKSVLYDENGCTFMKHQIGYDLHTPNVDS-----EENPDELFD 54
QY 58 CSCSRNMLTLADYCAQSVAGIGFDATC-SLYVLDKNGDPLPVSPEGDAKQNIYWMHRA 116
DB 55 LMTIKYVRESKVKDDIKFVSFSQMSHSLAMDQOHR-----TNNITWADNR 105
QY 117 TEQERINATH--HPVLNVGKISPEMETPKILMKENPEIYERAGOFPLADFLTW 174
DB 106 AKYATVINEVHDGNAIYQRTGPIHPMSPLAKIFMKHEMDVQRTAKAFDITTYIF 165
QY 175 A--TGDLARVCYVTCWKWTLAHENRWDPDYFRTIGLA-ELADEDFIRIGHIIVSPG 231
DB 166 LFDYVLIIDYSMASATGMFNLETLID--MDVEALELIGISKEMLP-----LVPT 214
QY 232 GNGTLAQAAMEMGLPGTPAVANGLIDAHAGIGTVGEGGLNNLAVFEGSSCTMAST 291
DB 215 MKGKERRATLMLGNKDTPEYIGASD--GVLNLSGVNSVKGGEVAVTIGTGAIRTV 271
QY 292 SP-----SFVPGVWGPPYYSAMVPGMLVVEGSGSAGAALDOLDLDFHRAV 335
DB 272 KPRIDYGRIFCYVLAEDHYIG--GPVNNGGVYLRML-----RDELL--ASEV 316
QY 336 EEARERMAORVNOPLPVLADRILEKTAQPSDAVALAKGLHVPFLGNRAFPADPHAR 395
DB 317 ETAKRLGV--DPYDVLQ--IAKRVKPG-----ADGLIFHPYLAGERAPLMMANAR 365
QY 396 ICGIGMERDLDNLALTYAGLCIGYGLROI-LDAQTAQGVNSKNIYISGAGGHPVLRQ 454
DB 366 FFGTLTSHKKEHMR--AALEGVLYNLYTYVLLIEVMNETPMIKATGGAFASEVW 422
QY 455 ILADTCGIPVITTOCEPVLGSAILGAVA-----GNIPASVEAMQOEFHNVK 504
DB 423 MMSDIFTELVPESYESSGACVGLKAVGDIEDFSIYSSMGATNNHPIEENTVY 482
QY 505 YFPOERQSLHHRREYAKOL-----OHTAK 530
DB 483 QETVSIPIINLSRSLTENYEQIADFOROHIAE 513
```

```

RESULT 6
US-09-815-242-5796
; Sequence 5796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
```

```

; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: us/09/815,242
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5796
; LENGTH: 509
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-5796
```

Query Match 10.2%; Score 288; DB 10; Length 509;

Best Local Similarity 23.6%; Pred. No. 3e-18; Matches 122; Conservative 83; Mismatches 221; Indels 90; Gaps 21;

```

QY 3 ITKTIVGVDSGVSRAIGFDLNGS-LLSHA-----TEKITTRSGSRVSSOEIMQ 55
DB 1 IMKYMIGVIGITSTKSVLYDENGCTFMKHQIGYDLHTPNVDS-----EENPDEL 53
QY 56 AVCSICRNALTLADYCAQSVAGIGFDATC-SLYVLDKNGDPLPVSPEGDAKQNIYWM 114
DB 54 AVLMTIKYVRESKVKDDIKFVSFSQMSHSLAMDQOHR-----TNNITWADN 104
QY 115 RATQERINATH--HPVLNVGKISPEMETPKILMKENPEIYERAGOFPLADFLT 172
DB 105 RAAKYATVINEVHDGNAIYQRTGPIHPMSPLAKIFMKHEMDVQRTAKAFDITTYIF 164
QY 173 WRA--TGDLARVCYVTCWKWTLAHENRWDPDYFRTIGLA-ELADEDFIRIGHIIVSP 229
DB 165 YHLEDFYLIIDYSMASATGMFNLETLID--MDVEALELIGISKEMLP-----LVPT 213
QY 230 PCGNGTLAQAAMEMGLPGTPAVANGLIDAHAGIGTVGEGGLNNLAVFEGSSCTMAS 289
DB 214 YVMKMKERRATLMLGNKDTPEYIGASD--GVLNLSGVNSVKGGEVAVTIGTGAIRTV 270
QY 290 TTSP-----SFVPGVWGPPYYSAMVPGMLVVEGSGSAGAALDOLDLDFH 333
DB 271 IDRRTDYGRIFCYVLAEDHYIG--GPVNNGGVYLRML-----RDELL--AS 315
QY 334 AVEARERMAORVNOPLPVLADRILEKTAQPSDAVALAKGLHVPFLGNRAFPADPHAR 393
DB 316 EETAKRLGV--DPYDVLQ--IAKRVKPG-----ADGLIFHPYLAGERAPLMMANAR 364
QY 394 AVTCGLMERDLDNLALTYAGLCIGYGLROI-LDAQTAQGVNSKNIYISGAGGHPV 452
DB 365 GSEFGTLTSHKKEHMR--AALEGVLYNLYTYVLLIEVMNETPMIKATGGAFASEV 421
QY 453 ROLLADTCGIPVITTOCEPVLGSAILGAVA--GNI 487
DB 422 ROMMSDIFTELVPESYESSGACVGLKAVGDI 457
```


Db 273 LMNTGEAVSENGSLTTACGPTGEVN---YALGEAVFAGASIO----- 315
QY 340 EMAORVNPPLPWLADRI-LEKTAOPSDAVAL---AKGLHVVEFLGNRAPFADPHARA 394
Db 316 -----MLRDEMKLINADYDSEYFATKYQNTNGVYVAPFGLGAPYNDPARG 363
QY 395 VIGGLGHERDLNLAALYIGLCIGYGLROIIDA-OTAGGVSKNIVISGAGOHPLVR 453
Db 364 AIFGLTRGVANNNHIIIR---ATLESIAVOTRDLVEAMQADSGIRLHALRVGGAANNFLM 420
QY 454 QIADTGGIVITTOCEPVLGSAIIGAVAGNIAPSVGEAMQOFTVDKYYP-----OE 509
Db 421 QFOSDILGTVEREVEYVATLGAAYLAGLAVGFQWNLDE-LOEKAVIEREFGIETTE 479
QY 510 RYOSLHRRRYEAYKQ 524
Db 480 R-----NYRYAGWK 489

RESULT 9
US-09-815-242-13431
; Sequence 13431, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13431

Query Match 9.8%; Score 276.5; DB 10; Length 502;
Best Local Similarity 24.4%; Pred. No. 3.4e-17;
Matches 140; Conservative 83; Mismatches 235; Indels 115; Gaps 23;

QY 1 MTTTIGVGVGSGVRAGIFDLNGSLSHATEKITTTTRSGSRVDSOSSOEIQAQASC 60
Db 1 MSQEKYIMADGGTTSSRAIIFNNKKGKSVSSQKEFTQIRPGAGVVEHNNNEIMNSQSV 60
QY 61 IRNALTLADYCAOSVAGIGF-DATCSLVLDKMGDPLVSPBGDAKONITVMDHRTAQ 119
Db 61 IAGAFIESGVKPNQIEAIGITNORETTVVWDKR-TGLPI-----YNATVMSRQTAPL 112
QY 120 AERINTNHHVNLVY-----GKISPEMETPKILMLKNNMPETVYRAQO----FPDLA 168

Db 113 AEOLKSO-----GYEFHEKTEGLIIDAYFSATKVRNIIIDHVEGAQRAEKELLFCTID 167
QY 169 DFLTRATGDLARSVCVTCKMTWLAH--ENRMPDPFRTIGLAEADEDFIRGHHIYS 226
Db 168 TWLVMKTLIDGAHAHTDVSNAARTMLYNIKELKMODELEIINIKALPE-VRSNSELV 225
QY 227 PGTFCGNGLTNOAAENKGLDGPVAVGLIDAHAGIGTVGVEGALNNLAVFGTSSCT 286
Db 226 -----GKTAPHFHYGGEVPISGMAC---DOQALFGOLAFEBGMWKN--TYGTSGFI 272
QY 287 MAST-----TSPSPVPGWMPYYSAMVGLIMLVGEGGSAGAATIDOLDHPA 334
Db 273 IMNTGEKMOSENLLTTIGT--GINKRYIAL-----EGSFTIGSALQMLRDLGRM 323
QY 335 VEEAREMAORVNPPLPWLADRILEKTAOPS---DAVALAKGLHVVEFLGNRAPFADPH 391
Db 324 VENSPE-----EKYARBSHNNDEV-----YVPAFGLGAPYNNQN 360
QY 392 ARAVTCGLGHERDLNLAALYIGLCIGYGLROIIDA-OTAGGVSKNIVISGGA 446
Db 361 ARGSVFGLTRGTSKEDFK--ATLOSIAVOVRDIDTMQVDTOTAIQVLK---VDGGA 413
QY 447 GQHPVLROIADTGGIVITTOCEPVLGSAIIGAVAGNIAPSVG-----EAMQOFTHV 501
Db 414 AMNFFLMQFOADIDIGIDIAKKNLETTALGAAFLAGL-----SVGYWKOLDLKLNET 467
QY 502 DKYYP--OERYOSLHRRRYEAYKQLOHTRAKL 531
Db 468 GELFEPMSNESRKRQLYKGMKAVKATQVFAEV 500

RESULT 10
US-09-815-242-13660
; Sequence 13660, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13660
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13660

Query Match 9.7%; Score 273.5; DB 10; Length 502;
Best Local Similarity 24.4%; Pred. No. 6.4e-17;


```
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5321

Query Match
Best Local Similarity 21.9%; Pred. No. 4.2e-16;
Matches 112; Conservative 77; Mismatches 227; Indels 95; Gaps 17;

OY 5 KTVIGDVSGSVRAGIFDLNGLSLSHATEKITTRRSRSGSVEQSSOEIMQAVCSIRNA 64
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 3 KYLISIOGTTSSRAILFNOKGEIAGVAQREFKQYFQSGVHEHDANEIMTSVLAVMTEV 62
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 65 LTLADVCAQSVAGIGF-DATCSLVYLDKN-GDPLVSPREGAKNIITVMDHATEQAER 122
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 63 INENDVRADQIAGIGITNORETIVVMDKHGTRPI-----YHAIWOSROTQSICSE 113
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 123 I--NATHNPVLNYGKISPEMETPKILMLKENNPEIYERAGQ----FFDLADFLTRAT 176
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 114 LKQOGEQTFERDKTGLLDLPFACTKYKWLIDNVEGAREKANGDLEFGTIDTWLVKLS 173
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 177 GDLARSVCTVCKWTWL--AHENRMDPYERTIG----LAELADEDFI--RIGHIIVS 226
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 174 GKAAHITDYSNASTLMFNHIDEMDELLELLTPKRNMLPEVAPSESVYKTIIDYHYG 233
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 227 PGTPCGN-----GLTAQAAEMGLLPGTPVAVGLIDAHAGIGITVGEAGALNNLAV 279
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 234 QEVPIAGVAGDQALRGACFERGDVKNYGTGFMIMNTGD--KAVKSEGLTTTAY- 291
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 280 FGTSSCTMASTSPSPVPGVWPYYSAMVPGMLVEGGSAGAAIDQLDHPHVEEAR 339
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 292 -----GIDGKYNVAL-----ESIFVSGSAIQ----- 313
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 340 EMAORVNPPLPWLADRI-LEKTAOPSDAVAL---AKGLHVEPELGNRAPDPHARA 394
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 314 -----WLRDGLRMINSAPQSESYATRVDSIEGYVVPFAFGLPYMDSEARG 361
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 395 VICGLGMRDLNLLALYIAGLCIGYGLROIIDAOTA--QGVSKNIVISGAGQHPVR 453
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 362 AIFGLTGTGTEKEHFIKATLESIC---YQTRDVMAMSKDSGIDVQSLRVDGAVKNNFIM 418
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 454 QILADTCGIPYITTCCEPVLLGSAILGAVA 484
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 419 QFOADIYNTSVERPEIQETTALGAVALGCLA 449
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 13
; Sequence 12335, Application US/09815242
; Patient No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12335
; LENGTH: 498
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12335

Query Match
Best Local Similarity 21.9%; Pred. No. 4.3e-16;
Matches 112; Conservative 77; Mismatches 227; Indels 95; Gaps 17;

OY 5 KTVIGDVSGSVRAGIFDLNGLSLSHATEKITTRRSRSGSVEQSSOEIMQAVCSIRNA 64
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 3 KYLISIOGTTSSRAILFNOKGEIAGVAQREFKQYFQSGVHEHDANEIMTSVLAVMTEV 62
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 65 LTLADVCAQSVAGIGF-DATCSLVYLDKN-GDPLVSPREGAKNIITVMDHATEQAER 122
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 63 INENDVRADQIAGIGITNORETIVVMDKHGTRPI-----YHAIWOSROTQSICSE 113
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 123 I--NATHNPVLNYGKISPEMETPKILMLKENNPEIYERAGQ----FFDLADFLTRAT 176
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 114 LKQOGEQTFERDKTGLLDLPFACTKYKWLIDNVEGAREKANGDLEFGTIDTWLVKLS 173
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 177 GDLARSVCTVCKWTWL--AHENRMDPYERTIG----LAELADEDFI--RIGHIIVS 226
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 174 GKAAHITDYSNASTLMFNHIDEMDELLELLTPKRNMLPEVAPSESVYKTIIDYHYG 233
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 227 PGTPCGN-----GLTAQAAEMGLLPGTPVAVGLIDAHAGIGITVGEAGALNNLAV 279
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 234 QEVPIAGVAGDQALRGACFERGDVKNYGTGFMIMNTGD--KAVKSEGLTTTAY- 291
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 280 FGTSSCTMASTSPSPVPGVWPYYSAMVPGMLVEGGSAGAAIDQLDHPHVEEAR 339
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 292 -----GIDGKYNVAL-----ESIFVSGSAIQ----- 313
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 340 EMAORVNPPLPWLADRI-LEKTAOPSDAVAL---AKGLHVEPELGNRAPDPHARA 394
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 314 -----WLRDGLRMINSAPQSESYATRVDSIEGYVVPFAFGLPYMDSEARG 361
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 395 VICGLGMRDLNLLALYIAGLCIGYGLROIIDAOTA--QGVSKNIVISGAGQHPVR 453
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 362 AIFGLTGTGTEKEHFIKATLESIC---YQTRDVMAMSKDSGIDVQSLRVDGAVKNNFIM 418
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 454 QILADTCGIPYITTCCEPVLLGSAILGAVA 484
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 419 QFOADIYNTSVERPEIQETTALGAVALGCLA 449
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 14
; Sequence 13788, Application US/09815242
; Patient No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```

? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 13788
? LENGTH: 501
? TYPE: PRT
? ORGANISM: Salmonella typhi
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: (1)...(501)
? OTHER INFORMATION: Xaa = Any Amino Acid
JS-09-815-242-13788

```

Db 437 PVMRETTALGAAYLAGLACGFWSLDE-LKSKAVIERVEPE 477

Search completed: March 13, 2003, 16:51:23
Job time : 17.0298 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 18.5662 Seconds
(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208b-4

Perfect score: 2811

Sequence: 1 MTITKTVIGVDVSGSGSVRAG.....HHRYEAYKQLQHTAKLLRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: pir1:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2031	72.3	535	2	S78598
2	1666	59.3	536	2	A99199
3	1666	59.3	536	2	AH3087
4	1560.5	55.5	538	2	AB3632
5	1244.5	44.3	545	2	AF0442
6	1167.5	41.5	569	2	D85354
7	971	34.5	525	2	A96183
8	971	34.5	525	2	A13103
9	864.5	30.8	715	2	S52675
10	465	16.5	509	2	A95985
11	448.5	16.0	560	2	D69587
12	405.5	14.4	566	2	C90637
13	405.5	14.4	566	2	C85488
14	402	14.3	563	2	H83883
15	394.5	14.0	566	1	KIECRU
16	381.5	13.6	500	2	D97221
17	377	13.4	569	2	AD0515
18	368.5	13.1	530	2	AF0051
19	368.5	13.1	569	1	A24984
20	362	12.9	492	2	C72417
21	361	12.8	567	2	AF0274
22	353	12.6	530	2	AE0941
23	352.5	12.5	542	2	S63222
24	340	12.1	545	2	B89823
25	336	12.0	493	2	G69468
26	334.5	11.9	502	2	E83353
27	328.5	11.7	504	2	AF1789
28	326	11.6	530	2	F90893
29	326	11.6	530	2	C85724

30	325	11.6	501	1	S18562	xyulokinase (EC 2
31	324	11.5	501	2	G75337	glycerol kinase -
32	323	11.5	530	2	B64905	sugar kinase homol
33	322.5	11.5	499	2	D69735	xyulose kinase xy
34	322.5	11.5	509	2	G95801	probable L-xyulok
35	321.5	11.4	499	2	S74045	glycerol kinase (E
36	320	11.4	504	2	AG1413	gluconate kinase h
37	319.5	11.4	511	2	G98252	cryptic L-xyulose
38	319	11.3	494	2	F83199	probable carboxydr
39	317.5	11.3	524	2	JN0606	ATP-stimulated glu
40	316	11.2	506	2	B72396	sugar kinase, FGgr
41	316	11.2	513	2	D72753	probable xyulose
42	315.5	11.2	492	2	AE3033	L-xyulose kinase
43	314.5	11.2	519	2	E83268	probable carboxydr
44	312.5	11.1	501	2	A99319	transporter/facili
45	310.5	11.0	524	1	S36175	glycerol kinase (E

ALIGNMENTS

RESULT 1
S78598
D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 17-Mar-2000
C:Accession: S78598; 508571
R:Heuvel, H.; Shakeri-Garakan, A.; Turgut, S.; Lengeler, J.W.
Submitted to the EMBL Data Library, January 1998
A:Description: Genes for D-arabitol and ribitol catabolism in enteric bacteria.
A:Reference number: S78598
A:Accession: S78598
A:Molecule type: DNA
A:Residues: 1-535 <HEU>
A:Cross-references: EMBL:AF045244
A:Experimental source: Strain KAT2026
R:Loviny, T.; Norton, P.M.; Hartley, B.S.
Biochem. J. 230, 579-585, 1985
A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural ge
A:Reference number: S07135; MIMD:86050424; PMID:2933028
A:Accession: S08571
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-16 <LOV>
A:Experimental source: strain FG5
C:Genetics:
A:Gene: rbtK
C:Function:
A:Description: catalyzes phosphorylation of D-ribulose to D-ribulose-5-phosphate
C:Superfamily: ribulokinase
C:Keywords: phosphotransferase

Query Match 72.3%; Score 2031; DB 2; Length 535;
Best Local Similarity 72.3%; Pred. No. 2.1e-144;
Matches 384; Conservative 49; Mismatches 98; Indels 0; Gaps 0;

OY	4	TKTVIGVDVSGSVRAGIIPDLNGLSHATEKITTTTFRSGSRVQSSQELTMOAVCSIRN	63
DB	5	TQNTIGVDVSGSVRAGVFNLRGELLAHATREITLFSAGNKVQSSREITMOAVCYCTIK	64
OY	64	ALTIADVCAOSVAGIGFDATCSLVLDKNGDPLVPSPEDAKONIIYVMDHRTAQER	123
DB	65	AVANAGVSPSSINGIGFDATCSLVVIGDNDAPLAVGSDADNRITVMDHRTAQER	124
OY	124	NATHHPVLYNVGKISPEMETPKILMLKEMPEITERRAGOFFDLADFLYTRATGDLARV	183
DB	125	NATGHVPLRVYVGKISPEMETPKILMLKEMRPHIYQLARHFFDLADFLYTRSGDEARV	184
OY	184	CYTTCCKWTVAHNRMDPQVFRITIGLAEDDFIRIGHIIVSPGTCGNGLTQAQAAEM	243
DB	185	CYTTCCKWTVAHNRMDVAGFFRQIGLELVDEDFVRIGRITVDPGPGCGGLCATAAEEM	244
OY	244	GLLPGRFVAVGLIDAHAGIGTGVVEGALNNLAVYVGTSSCTMASTSPSPFVGVWGPY	303

```
Db 245 GLPIGTAAAGMDAHAGGIGTGTGVLNGAVNNMAYVFGTSSCTMTTQOEAFFVGVWGPY 304
QY 304 YSAMVPEGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQ 363
Db 305 YSAMVPEGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQ 364
QY 364 PSDAVALAKGLHVVPEFLGNRAPRAPPDHPARAVICGLMERDLDNLALYAGLCIGYGL 423
Db 365 PSEAVTLAAGLHVVPEFLGNRAPRAPPDHPARAVICGLMERDLDNLALYAGLCIGYGL 424
QY 424 RQIDAQTAQGVSKNIVISGAGQHPVLVQIADTGCIPVITTOCCPEVLLSGAIIIGAV 483
Db 425 RQIDAQTAQGVSKNIVISGAGQHPVLVQIADTGCIPVITTOCCPEVLLSGAIIIGAV 484
QY 484 AGNIAPVGEAMQOFTVVDKYYVQERQSLHRRYEAYKOLQHTAKLRD 534
Db 485 AGRAVLAALPEAMKQFTQVADATYHSETAFSPULHQRVYAAKALQOAGRLIRE 535
```

RESULT 2

D:ribulokinase (EC 2.7.1.47) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: A99199
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: A99199
A:Gene: AGL_L1075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89115.1; PID:g15158921; GSPDB:GN00170
C:Genetics:
A:Map position: linear chromosome
C:Superfamily: ribulokinase
C:Keywords: phosphotransferase

Query Match 59.3% Score 1666; DB 2; Length 536;
Best Local Similarity 60.0%; Pred. No. 4.9e-117; Mismatches 146; Indels 0; Gaps 0;
Matches 316; Conservative 65; Mismatches 146; Indels 0; Gaps 0;
QY 7 VIGDVGSGVRAIGIFDLNGLSLSHATEKITTTTTRSGRVSQSOEIQAVCSIRNALT 66
Db 5 LVGVDTGTSARAGVFDVAGKLTAKRPISMHRDEGIAQSSAEVWQAVCDVSQSVS 64
QY 67 LADVCAQSVAGIGDATCSLVYLDKNGDPLVPSPEGDAKONIYWMHRAVEQAEERINAT 126
Db 65 RAGIDPAEVNTIGFDATCSLVVRGPGDVTLPVGAADHPERDIYWMHRAVEQAEERINAG 124
QY 127 HHPLVNVVGGKISPEMETPKILMKENMPETIERAGOFDLADFLTPVATGDLARSCTV 186
Db 125 KHAVLKTVGGKISPEMETPKILMKENMPETIERAGOFDLADFLTPVATGDLARSCTV 184
QY 187 TCKMTWLAHENRMDPDYFRTIGLAELADEDEFIRIGHIIVSPGTCNGLTAAQAAEMGL 246
Db 185 TCKMTWLAHENRMDPDYFRTIGLAELADEDEFIRIGHIIVSPGTCNGLTAAQAAEMGL 244
QY 247 PGTVAAVGLIDAHAGGIGTVEGAGLNNLALYVGTSSCTMASTTSFVGVWGPYISA 306
Db 245 AGTAAVAGLLIDAHAGGIGTVEGAGLNNLALYVGTSSCTMASTTSFVGVWGPYISA 304
QY 307 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQPSD 366
Db 305 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQPSD 364
QY 367 AVAALAKGLHVVPEFLGNRAPRAPPDHPARAVICGLMERDLDNLALYAGLCIGYGLRQI 426
Db 365 AVAALAKGLHVVPEFLGNRAPRAPPDHPARAVICGLMERDLDNLALYAGLCIGYGLRQI 424

```
QY 427 LDAQTAQGVSKNIVISGAGQHPVLVQIADTGCIPVITTOCCPEVLLSGAIIIGAV 486
Db 425 IETQARNCAPVETISVSGAGAHPLARQLADATGCLPELTCEPEVLLSGAIIIGAV 484
QY 487 IAPVGEAMQOFTVVDKYYVQERQSLHRRYEAYKOLQHTAKLR 533
Db 485 TYPLMAMAPMSRIDSSAMPDPDFOKIHOARVDAFLALQNAARAIR 531
```

RESULT 3

ribulokinase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH3087
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woode, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; McCl, Karp, P.; Romero, P.; Zhang, S.; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH3087
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <KUR>
A:Cross-references: GB:AE00689; PIDN:AL45118.1; PID:g17742788; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4324
A:Map position: linear chromosome
C:Superfamily: ribulokinase

Query Match 59.3% Score 1666; DB 2; Length 536;
Best Local Similarity 60.0%; Pred. No. 4.9e-117; Mismatches 146; Indels 0; Gaps 0;
Matches 316; Conservative 65; Mismatches 146; Indels 0; Gaps 0;
QY 7 VIGDVGSGVRAIGIFDLNGLSLSHATEKITTTTTRSGRVSQSOEIQAVCSIRNALT 66
Db 5 LVGVDTGTSARAGVFDVAGKLTAKRPISMHRDEGIAQSSAEVWQAVCDVSQSVS 64
QY 67 LADVCAQSVAGIGDATCSLVYLDKNGDPLVPSPEGDAKONIYWMHRAVEQAEERINAT 126
Db 65 RAGIDPAEVNTIGFDATCSLVVRGPGDVTLPVGAADHPERDIYWMHRAVEQAEERINAG 124
QY 127 HHPLVNVVGGKISPEMETPKILMKENMPETIERAGOFDLADFLTPVATGDLARSCTV 186
Db 125 KHAVLKTVGGKISPEMETPKILMKENMPETIERAGOFDLADFLTPVATGDLARSCTV 184
QY 187 TCKMTWLAHENRMDPDYFRTIGLAELADEDEFIRIGHIIVSPGTCNGLTAAQAAEMGL 246
Db 185 TCKMTWLAHENRMDPDYFRTIGLAELADEDEFIRIGHIIVSPGTCNGLTAAQAAEMGL 244
QY 247 PGTVAAVGLIDAHAGGIGTVEGAGLNNLALYVGTSSCTMASTTSFVGVWGPYISA 306
Db 245 AGTAAVAGLLIDAHAGGIGTVEGAGLNNLALYVGTSSCTMASTTSFVGVWGPYISA 304
QY 307 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQPSD 366
Db 305 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNOPLPWLADRIIEKTAQPSD 364
QY 427 LDAQTAQGVSKNIVISGAGQHPVLVQIADTGCIPVITTOCCPEVLLSGAIIIGAV 486
Db 425 IETQARNCAPVETISVSGAGAHPLARQLADATGCLPELTCEPEVLLSGAIIIGAV 484
QY 487 IAPVGEAMQOFTVVDKYYVQERQSLHRRYEAYKOLQHTAKLR 533

Db 485 TYPDLMAMPAMSRIDSSAMPDPDFOKIHOARYDAFLALONARAIR 531

RESULT 4

AB3632

D:ribulokinase (EC 2.7.1.47) [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 17-May-2002

C:Accession: AB3632

R:DeVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Jos, T.; Ivanova,

; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AB3632

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-538 <KUR>

A:Cross-References: GB:AB008918; PIDN:AAL54221.1; PID:g17985192; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME10979

A:Map position: 11

C:Superfamily: ribulokinase

C:Keywords: phosphotransferase

Query Match 55.5%; Score 1560.5; DB 2; Length 538;
Best Local Similarity 57.8%; Pred. No. 4e-109;

Matches 309; Conservative 66; Mismatches 155; Indels 5; Gaps 3;

2 TTTKTVIGDVGSVRAIGFDLNGSLSHATEKITTTTTRSGSRVEOSSQEIWOAVCSCT 61

4 TMTNYLVGVDVGTSGARAGLFDAGGTMLASARDIAIMWEGAGIVGSSDDIMQANCEV 63

62 RNALFLADVCAOSVAGIGCDATCSLVLDKNDPLVSPEDGAKONIYWMHRAEQAE 121

64 REVVAVAGDPAVAGIGCDATCSLVLDKNDPLVSPEDGAKONIYWMHRAEQAE 123

122 RNTNHHPLVAVGKISPEMETPKILMKEMPEIYERAGOFPLADLTTRATGDLAR 181

124 RINTTKADVLGVGAGISEMETPKILMKEMPEIYERAGOFPLADLTTRATGDLAR 183

182 SVCTVTKMTWLAHNRMDPDYFRTIGLAELADEDFIRIGHIVSPGTCNGITPAQAA 241

184 SACTVTKMTWLAHNRMDPDYFRTIGLAELADEDFIRIGHIVSPGTCNGITPAQAA 242

242 EMGLLPGRVAVGLIDAHAGIGTVGV---EGALNINLAIVFGTSSCTMASTTSPSEVPG 298

243 ELGLRPGTAIAAGLIDAHAGIGTVGV---EGALNINLAIVFGTSSCTMASTTSPSEVPG 302

299 VNGPYTSANVPGLMLVEGOSAGAAIDQLDFHRAVEAREMAQRVNOPPLVWLADRL 358

303 VNGPYTSANVPGLMLVEGOSAGAAIDQLDFHRAVEAREMAQRVNOPPLVWLADRL 361

359 ERTAOPSDAVNALAGLHVHPEFLGNRAPADPHARAVICGLMERDNLMLALYIAGLCG 418

362 EARGGPEKMTAVIGDIHVHPEFLGNRAPADPHARAVICGLMERDNLMLALYIAGLCG 421

419 IGYGLRQILDAGTAAOVVSKNIVISGAGOHPLVROLADTGCIPVITTOCEPVLGSA 478

422 LGYGRQIIEADRAKGIIVDTIVVSGAARSMLVROVLADACGLVYTAATSEPEVYLGA 481

479 ILGAVAGNIAPSVGEAMQOFTHVDKYRQERYQSLHHRREYAKQLOHTAKLR 533

482 MIGAASGAYPDLVTAMOVMSSELGARNRDPARRAKMHDHREFAFMLOQTARKIR 536

RESULT 5

AF0442

Probable carbohydrate kinase YPO3637 [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AF0442

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tildall, R.W.; Holden, M.T.G.; Prentice, M

demo-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0442

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <KUR>

A:Cross-References: GB:AL590842; PIDN:CAC93106.1; PID:g1598158; GSPDB:GN00175

C:Genetics:

A:Gene: YPO3637

A:Map position: 4

Query Match 44.3%; Score 1244.5; DB 2; Length 545;
Best Local Similarity 45.3%; Pred. No. 2e-85;

Matches 241; Conservative 100; Mismatches 174; Indels 17; Gaps 5;

8 IGVVGSVRAIGFDLNGSLSHATEKITTTTTRSGSRVEOSSQEIWOAVCSCTRNA 67

6 IGVVGSVRAIGFDLNGSLSHATEKITTTTTRSGSRVEOSSQEIWOAVCSCTRNA 65

68 ADVCAOSVAGIGCDATCSLVLDKNDPLVSPEDGAKONIYWMHRAEQAE 127

66 ADVCAOSVAGIGCDATCSLVLDKNDPLVSPEDGAKONIYWMHRAEQAE 125

128 HPLVAVGKISPEMETPKILMKEMPEIYERAGOFPLADLTTRATGDLAR 187

126 HPLVAVGKISPEMETPKILMKEMPEIYERAGOFPLADLTTRATGDLAR 185

188 CKMTWLAHNRMDPDYFRTIGLAELADEDFIRIGHIVSPGTCNGITPAQAA 247

186 CKMTWLAHNRMDPDYFRTIGLAELADEDFIRIGHIVSPGTCNGITPAQAA 245

248 GTPVAVGLIDAHAGIGTVGV---EGALNINLAIVFGTSSCTMASTTSPSEVPG 301

246 GTPVAVGLIDAHAGIGTVGV---EGALNINLAIVFGTSSCTMASTTSPSEVPG 305

302 PYTSANVPGLMLVEGOSAGAAIDQLDFHRAVEAREMAQRVNOPPLVWLADRL 357

306 PYTSANVPGLMLVEGOSAGAAIDQLDFHRAVEAREMAQRVNOPPLVWLADRL 360

358 LEKTA-OPSDAVNALAGLHVHPEFLGNRAPADPHARAVICGLMERDNLMLALYIAG 416

361 LEKTA-OPSDAVNALAGLHVHPEFLGNRAPADPHARAVICGLMERDNLMLALYIAG 420

417 CGIGYGLRQILDAGTAAOVVSKNIVISGAGOHPLVROLADTGCIPVITTOCEPVLG 476

421 GMLALGTRHILIEFMNONGNIDTMMASGGCTKPIVVEAHANATGCAMLLPESEAMLLG 480

477 SALGAVAGNIAPSVGEAMQOFTHVDKYRQERYQSLHHRREYAKQLOHTAKLR 527

481 SAMGTVAGVVESELPEDAMAMSRIGKTVTPQNKIKAYDRKRVVHOMYH 532

RESULT 6

DB5354

Hypothetical protein AT4g30310 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: DB5354

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: AB5001; MUID:20083488; PMID:10617198

A:Accession: DB5354

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-569 <STO>

A:Cross-References: GB:NC_001268; MUID:g7269931; PIDN:CA81024.1; GSPDB:GN00140

A:Gene: AT4g30310

A:Map position: 4


```

Db 124 ASGRHVLDPAGNSVSEPMOKPLMKLTMPQSWSRMSRFAFDLDELTKATCSAQRNC 183
Oy 185 TYTCKMTWLAHEN-RWMDPYFRITGLAEADEDFTRIG--HHIYSPGTCNGSLTQAAA 241
Db 184 TOTAKMNFIAQHPGQADYLAFLADLDKE-----RAGLPETTMGGAIG-PLSEAAA 238
Oy 242 EMGLLPGTPPAVGLIDAHAGIGTGVESGAL-----NNLAVVFCSCTMASTTSPFV 296
Db 239 ELGLDPTCCVAAQMDIAYAGALGAL--GGCLAEQYGRKVALIAGTSSCLVAMSTP--M 293
Oy 297 PG--VMGPRYSAMVPEGLMVEGQSAAGAIQDQLDFHRAVEBAREMAORVNPFLPWL 354
Db 294 PERSLIMGPRWQAVLPGHMLVEGQSATGALLDHIYRMHAAGEP-----DTALHARIV 346
Oy 355 DILLEKTQPSDAVALAKLHVPELGNRAPADPHAAVYICGLMERDLDNLALYTA 414
Db 347 AVNTE--LNELEGEAFADLHVLPDFHGRSPDLADPHAAVSGTLTDFSPSLCYLNR 404
Oy 415 GLCGICGYCRLQILDQTAQGVVSKNIVISGAGOHPLVQIADTCIPVITTOCEPYL 474
Db 405 TPAIALAGRHVLDMERGVAVETLHYGVHKNPLMELADYTKRIYVATADAVL 464
Oy 475 LGSAILGAVAGNIADSV--GEAMQO-----FTHVDKYYYRPOERYOSLHRRYE 520
Db 465 LGTAMTAATAGCVHASLAAGAAMYPGNAELSGNPALAHYERDY---RRFLAMYRHRQ 521

```

RESULT 9

Probable membrane protein YDR109c - yeast (*Saccharomyces cerevisiae*)
 N.Alternate names: hypothetical protein YD9727.05c
 C.Species: *Saccharomyces cerevisiae*
 C.Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 C.Accession: S52675
 R.Murphy, L.; Shore, L.; Harris, D.
 Submitted to the EMBL Data Library, March 1995
 A.Reference number: S52671
 A.Accession: S52675
 A.Molecule type: DNA
 A.Residues: 1-715 <MUR>
 A.Cross-references: EMBL:Z48758; NID:g747879; PID:g747884; GSPDB:GN00004; MIPS:YDR109c
 C.Genetics:
 A:Gene: MIPS:YDR109c
 A:Cross-references: SGD:S0002516
 A:Map position: 4R
 C.Keywords: transmembrane protein
 F:515-531/Domain: transmembrane #status predicted <TM1>
 F:535-551/Domain: transmembrane #status predicted <TM2>

```

Query Match          30.8%; Score 864.5; DB 2; Length 715;
Best Local Similarity 36.1%; Pred. No. 9, 1e-57;
Matches 201; Conservative 100; Mismatches 211; Indels 45; Gaps 10;

Oy 5 KTVIGVDVSGSVRACIFDLNGLSLSHATEKITTTTRSGSRVQSSQEIWOAVCSIRNA 64
Db 38 KYVGVADVGTGARACVIDOSGNMLSLAKRPKIREQLISNFTQSSREIMNAVCCVRRY 97
Oy 65 LTLADYCAQSVAGIGFDATCSLVLD-KNGDPLPVSP- GDAKONITVMDHATQDAER 122
Db 98 VEESGVDPFRVRGIGFDATCSLVVVSATNFEELAVGPDPDTNNDONITLMDHAMEETEE 157
Oy 123 INATHHPVLNVYGGKISPEMETPKILMLKENMPEIYERAGOFDLDLFTWRATGDLARS 182
Db 158 INSSGCKCLKYVGGQSVEMETPKIKMLKNLNLNLEAGIFQDCKFFLDLYLTKATGKENS 217
Oy 183 VCTVTCCKWTWL-----AHENRMDPDYFRITGLAEADEDFTRIG-----HHIYSPGTP 231
Db 218 PCSAVCKQGFPLPYGVGSDIGMSKEFLNSIGSELTKNDFERLQSGSLREKKNFLTAG-EC 276
Oy 232 GNGCTQAAAENKGLPRTYVAVGLIDAHAGIGTGVV-----GVAL 273
Db 277 ISPLDKKAACQGLTEHCYVSSGIIDAYAGWGTVAKPESAVKGLAETENYKKDPNGAI 336

```

```

Oy 274 NNLAVVFGTSSCTMASTTSPFVPGVYSAVPEGLMVEGQSAAGAIQDQLDFH 333
Db 337 GRILAAVAGSTCHLILLSKNPFIYHGVWGPYRDLAAGPMAAEGGCGCTGLDHLTTFH 396
Oy 334 AVEAREMAORVNPFLPWLADRIELKTYAP--SDAVLAKLHVPEFLGNRAPADP 390
Db 397 AFEELSHMANLGVSRFEYL-NKILETLEVKRKRVSIVISLAKHLFPYGYDHGRNSPIADP 455
Oy 391 HRAVYICGLMERDLDNLALYTAGLCIGYIGRIQLDQTAQGVVSKNIVISGAGQHR 450
Db 456 NMRACITIGSMDSIEDLAVMLYLSACEFISQOTRILLEYWLSKGEHINALEMSGGCGRNS 515
Oy 451 LVROIADPTCGIPVITTOCE--PVLGSAITLGAACNIAPSVGEAMQOFTHVKKYYPQ 509
Db 516 LMRLLADCTGPIVIRYVDAVAVGFSALILGAASEDF-----DTRKRTLKQK 567
Oy 510 RQSLHRRYEAYKQD 526
Db 568 SSQTKTERFNDYSYSIQ 584

```

RESULT 10

probable sugar kinase protein (EC 2.7.1.-) [imported] - *Sinorhizobium meliloti* (strain C:Species: *Sinorhizobium meliloti*)
 C.Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C.Accession: A95985
 R.Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her
 A>Title: The complete sequence of the 1,683-kb pSYM megaplasmid from the N₂-fixing e
 A.Reference number: A95842; MUID:21396508; PMID:11481431
 A.Accession: A95985
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-509 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49545.1; PID:g15141032; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R.Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hu
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela
 hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
 A>Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A.Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Smb20852
 A:Genome: Smb20852
 C:Keywords: phosphotransferase

```

Query Match          16.5%; Score 465; DB 2; Length 509;
Best Local Similarity 26.3%; Pred. No. 5, 2e-27;
Matches 133; Conservative 89; Mismatches 213; Indels 70; Gaps 16;

Oy 7 VIGVDVSGSVRACIFDL-NGSLSHATEKITTTTRSGSRVQSSQEIWOAVCSIRNA 65
Db 4 VLSLDGTGGARAFDTQTNITVARGEARPYKQHLRPNRAEDNPEEDMTALVSLVPDY 63
Oy 66 TLA--DYCAQSVAGIGFDATCSLVLDKNGDPLPVSPEDAKONITVMDHATQDAER 122
Db 64 AKAGSPDIAAVCATF-----ASTVVLCDRSK--PIAP-----AVLMDARAADAAAF 110
Oy 123 INATHHPVLNVYGGKISPEMETPKILMLKENMPEIYERAGOFDLDLFTWRATGDLARS 182
Db 111 TETVDHPRLADSGSDAVEVLVPAKMFARPKRDLARREVLEALDFVNHRLTGWAGS 170
Oy 183 VCTVTCCKMTWLAHENRMDPDYFRITGLAELA--DEDETRIGHNIYSPGTCNGTLAQA 239
Db 171 LMAATCKMWDKSNRKFCELDYALFGVPDGLAALPQRIYDQ-DVAAPMLP-----EM 222
Oy 240 AEMGLLPGTPPAV-GLIDAHAGIGTGVGEGALNNLAVFGTSSCTMASTTSPS-FVP 297

```

223 ACTTG-TCGPNPVVGGIDAH---MGTEGDAVTPGSMFLFGTGSNVHLTQVPRDDGRNR 278

228 GWMGPRYTSAMPGLMVEGSGSAGACAIQDLDFHPAVEAREMAORVNOPLPWADRI 357

229 GWMGPRYTSAMPGLMVEGSGSAGACAIQDLDFHPAVEAREMAORVNOPLPWADRI 357

279 GWMGPRYTSAMPGLMVEGSGSAGACAIQDLDFHPAVEAREMAORVNOPLPWADRI 357

358 -----LEKTAQPSDAVAL-AGLHVVPFLGNRPAPADPARAVIGGIGMERDNL 409

314 FELDAGCRSLCAADALIEPDSGTGLALDYWMGNTPTRDARLKAFLGLSLSHP---RA 370

410 ALYTAGLCIGYGLRQIILDAQTAQGVSKNIVISGAGOHPLVROILADTCGPIVITQC 469

371 STYRAATVAVALGAAVNFDELEKGVCAIDRIIMSGIMKNRMLEATIDAIKPEVELALD 430

470 CEPVLGSLNLCAGVAGNIAPSVGEA 494

431 DNLSTHGAAGVACTVALGLFPLDTTA 455

RESULT 11

D69587

Ltritoluclease arab - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: D69587

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertea, C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chel, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A:Authors: Foulger, C.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis, A.; Lauther, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, Y.M.; Ogawa, K.; Ogigawa, A.; Oudgaa, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet, M.; Rivoita, C.; Roche, E.; Roche, B.; Rose, M.; Sadleir, V.; Sato, T.; Scanlon, T.; Schuch, M.; Tamakoshi, A.; Tanaka, T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winers, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A.; Yoshida, H.; Yoshikawa, H.; Yumoto, E.; Yoshikawa, H.; Yanchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: D69587

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-560 <KUN>

A:Cross-references: GB:259118; GB:AL009136; NID:g2635200; PIDN:CAB14839.1; PID:g2635344

A:Experimental source: strain 168

C:Genetics:

A:Gene: arab

C:Superfamily: rfbuloklase

Query Match 16.0%; Score 448.5; DB 2; Length 560;

Best Local Similarity 27.1%; Pred. NO. 1e-25;

Matches 153; Conservative 95; Mismatches 237; Indels 79; Gaps 20;

QY 8 IGVNDCGSGVRAGFDL-NGSLLSHATEKI-----TTTRSGSRYE-----QSSQETWQ 55

DB 5 IGVNDCGSGVRAGFDL-NGSLLSHATEKI-----TTTRSGSRYE-----QSSQETWQ 55

QY 56 AVSCGIRALTLADVCASVAGIGFDAT-CSIYVLDKNGDRLPVSPEDAKON--IYWM 112

DB 65 VLETTIPILLOTGVDPRDIIIGIGIDFTACTTLPIDSSGQRCMLPEYEEBNSYVUKMK 124

QY 113 DHRATEQAEKIN---ATNHPVNLVVGKISPEMETPKILMKEMPREYERAGOFPLA 168

DB 125 HHAAOKNADRLNOLAEESGEAFLORYGKISSEMMIPVMOIAEAPRITAEADRIEAA 184

QY 169 DELTWRATGDLARSYCTVTKMTWLAHNRMDPRDFRIGLA-ELADDFRIGHNIYSP 227

DB 185 DWIYVYOLGSLKRSNCTAGYKAMSEKAGYPSDDFEKLNPSMKITRD--KLSSHSIV 242

QY 228 STPGCNGITLQAQAAEMGLPSTPVAVALGLIDAHAGGIGTVG-EGGALNNLAVFGTSCT 286

Db 2A3 GEEKS -LTEKMAKLTGLPRAVANAVANDAHV -SVPAVGTIEG- ---KMLMINGTSTCH 297

Oy 287 MASTTSPSPVQVWGPVYSAWPGMLVEGQSAAGAAIDOLD- -FHPAV-EEARE- --- 340

Db 298 VLLGEVHIVVGMCGVNDGILLPGYAGYEGQSCGDHFDVFTCVCPAYQAEKKEKNI 357

Oy 341 -----MAORVQPLPVMILADRIEKTAPSDAVALAKLHVPEELGRAPFAOPHARAV 395

Db 358 GVHELLSKRANQAG- -----ESGLLDMWNGNRSTLVVDLTGM 398

Oy 396 ICGLMERDLDNLNLAYLGLGIGYGLROIIDAQTAQVYAKNVISGG-AGQHPVYQ 454

Db 399 LLGMLT---LTKPEIYALVATATGTRMIETPKESGVIEIELFPAAGLAENKPFVQ 455

Oy 455 ILADTCGIPVITTOCCPEVLGSAITLGAAG- -----NIAPSVGEAMQOETHDK 503

Db 456 IYADVTNMDDIKSGSPQAPALGATFGALAGKEKGYDIDIKKAANMGLK- ---DI 509

Oy 504 YVYPERYQSLHHRRYEAYKOLQ 527

Db 510 TYTPNAENAAVEKLYAEKELVH 533

RESULT 12

C90637

L-ribulokinase [imported] - Escherichia coli (strain O157:H7, substrain R1MD 050952)

C:Species: Escherichia coli

C:Accession: C90637

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

gasaRes. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome analysis of a closely related strain

A:Reference number: A99629; PMID:11558231; PMID:11558796

A:Accession: C90637

A:status: preliminary

A:molecule type: DNA

A:Residues: 1-566 <HAV>

A:Cross-references: GB:BA000007; PIDN:BA83490.1; PID:G13359523; GSPDB:G000154

A:Experimental source: strain O157:H7, substrain R1MD 050952

C:Genetics:

A:Gene: ECs0067

C:Superfamily: ribulokinase

Query Match 14.4%, Score 405.5; DB 2; Length 566;

Best Local Similarity 27.7%, Pred. No. 1.7e-22;

Matches 163; Conservative 77; Mismatches 224; Indels 125; Gaps 29;

Oy 8 IGVNNGSSVRAAGTDLNGLSLSHATEKITTTTRNSGRVDSQETWO-AYVSCIRNAL 65

Db 5 IGLDFGSDSVRALAVD-----CTTGEEIATSVEMVPR--WQGGQFCADPNQF 50

Oy 66 -----TLADYCAQ---SVAGIGFDATCSL-VYLDKKGDDLPVSPGDA 104

Db 51 RHHPRDYIESMEALKTYLAEISVEQRAAVVIGVDSGSPAPADAGNVALRPEAE 110

Oy 105 KQN---IIVMMDHRAEQEAERINATHHPVLN-----YVGGKISPMEETPKILMLKENMEPI 157

Db 111 NPNMAFVLMKQHTVAEEAEELTRLCHARGVNDYSYIGIYSSSEMFAKILHIVRQDSAV 170

Oy 158 YERAGQFDDLADFLTWARG-----DLARSVCTVTCKWTLAHENRMDPDYFRTIGL--A 210

Db 171 AQSASWIELCDWVALLSGTTRPQDIRGRCSAGHKSLM--HES-WQ-----GLPPA 220

Oy 211 ELADEDPRIRIGHIYSP-----GTPCGNGLITRQAQAAEGLRGPFPVAVGLIDAHAGG 262

Db 221 SFFELDLIRLKHRLPSPLEDTWTADIPVGT-LCEPMAQRIGLPESSVVISGARDCHMGA 279

Oy 263 IGTGVCEGALNNLILAYVYSSCTMASTTSPF-----VPGWVGYYASAMVAGLWLVGQ 318

Db 280 VGA-----GAQGNALYKVYGTSTCDILLADKQSVGERAVKVGICGQYDGSVYVGFGLGEGQ 335

Oy 319 SA-----AGAAIDOLDLPHPAVEAREMAQORVQPLP-----VMLADRIEKTAAQ 363

Db 336 SAFGDIYAMFGVRLGMPLEQLAQAHEPELKQIDASQK--QLPALTEAMAKNSLDH--- 390
QY 364 PDAVALALAGLHYVPE-FLGNRAPFADPHARAVICGLGMRDLDNLALYIAGLCIGYG 422
Db 391 -----LPVLLDMFNGRRTPMANQRLKGYITDLNLTADAPLLFGGLIA---ATAFG 437
QY 423 LRQILDQAOTAGVSKNIVISG-AGOHPLVRQILDATGCP---VITTOCCPEVLLGSA 478
Db 438 ARAIMCEFTDQGIANNVNMALGGIARKNOVIYMACCDVLRPLQIYASDQCC---ALGAA 494
QY 479 ILGAVAGNIAPSVGEAMQOF-THVDKYYYPQ---ERYOSLHRRREAY 522
Db 495 IFAVAIAKAVHADIPSAOKMASAVEKTIQPRSDQAFQEQ-LYRRYQOW 542

RESULT 13

C85488
L-ribulokinase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85488
Riperna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimlantia, E.; Potamocousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <STO>
A:Cross-references: GB:AE005174; NID:g12512760; PIDN:ACG54367.1; GSPDB:GND0145; UWGP:200
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: arab
C:Superfamily: ribulokinase

Query Match 14.4%; Score 405.5; DB 2; Length 566;
Best Local Similarity 27.7%; Pred. No. 1.7e-22;
Matches 163; Conservative 77; Mismatches 224; Indels 125; Gaps 29;

QY 8 IGVDVSGSVYRAGIFDLNGLSLSHATEKITTRRSGRVBOSSQEIWO--AVCSIRNAL 65
Db 5 IGLDFGSDSVRALVD-----CTTGEEIATSVEMYP--WQKQCFDAPNQG 50
QY 66 -----TLADVCAQ---SYAGIGFATGCL-VYLDKNGRPLVPSPEGDA 104
Db 51 RHHRPDYIESMEALKTVALEISVEQRAAVVIGVDSGTSTPAPIDADGNVLAIRPEFAE 110
QY 105 KON-IIVMDHRAEQERINATHPVLN-----YVGGKISPEMETPKILMKENPEI 157
Db 111 NPRAVMVLKMDHTAVEAEETITLCHAPGVNDYSRVIGGYSSEMFMAKILHTRQDSAV 170
QY 158 YERAGQFPLDLFLWBRATG-----DLARSVCVTCTKWTVAHNRMDPYFTTGL--A 210
Db 171 AQAAMMIEICDMVPPALLSGTTRPDRRCRCSAGHKSIL--HES-WG-----GLPPA 220
QY 211 ELADEDPIRGIHIVSP-----GTPCGNGILTAQAAAMGILLPGPRAVGLIDAHAGG 262
Db 221 SPFDELDPILRHLPSLTDTYADIPVGT-LCPREMAQRLGIPRESVVIGGAFDCMGA 279
QY 263 IGTVEGEGALNMLAVVFGTSSCTMASTTSPSF---VPGVMPYYSAMVPGMLVVEGQ 318
Db 280 VCA-----GAQPNALVYKIGISTCDILADKQSVGERAVKICQGVDSVYVPGFGLEAG 335
QY 319 SA-----AGAAIDQLDFHRAVEAREMAQVNPPL---VWLADRILEKTAQ 363
Db 336 SAEGDIYAMFGVRLGMPLEQLAQAHEPELKQIDASQK--QLPALTEAMAKNSLDH--- 390
QY 364 PDAVALALAGLHYVPE-FLGNRAPFADPHARAVICGLGMRDLDNLALYIAGLCIGYG 422
Db 391 -----LPVLLDMFNGRRTPMANQRLKGYITDLNLTADAPLLFGGLIA---ATAFG 437

QY 423 LRQILDQAOTAGVSKNIVISG-AGOHPLVRQILDATGCP---VITTOCCPEVLLGSA 478
Db 438 ARAIMCEFTDQGIANNVNMALGGIARKNOVIYMACCDVLRPLQIYASDQCC---ALGAA 494
QY 479 ILGAVAGNIAPSVGEAMQOF-THVDKYYYPQ---ERYOSLHRRREAY 522
Db 495 IFAVAIAKAVHADIPSAOKMASAVEKTIQPRSDQAFQEQ-LYRRYQOW 542

RESULT 14

H83883
L-ribulokinase arab [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: H83883
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; I
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans e
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: H83883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-563 <STO>
A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05591.1; GSPDB:C
A:Experimental source: strain C-125
C:Genetics:
A:Gene: arab
C:Superfamily: ribulokinase

Query Match 14.3%; Score 402; DB 2; Length 563;
Best Local Similarity 26.7%; Pred. No. 3.1e-22;
Matches 155; Conservative 82; Mismatches 260; Indels 84; Gaps 20;

QY 4 TKYIVGVDSGSYRAGIFDL-NG-SLSHATE-----KITTRRSGRVBOSSQ 51
Db 3 TKYITGVDSGSGRAVLIDLNSQELADHVTPRHGVIDOYLPTNTIKLGHEMALQHP 62
QY 52 EIMQAVCSIRNALTLADVCAQSVAGIGFAT-CSVYLDKNGRPLVPSPEGDAKON--- 107
Db 63 DYVEVLITSPRAYMKESGVADVDYIGVDFCTACTMLPVDEGQPLCLLD--YKDNPHS 120
QY 108 -IIVMDHRAEQERIN---ATNHPVLNVGKISPEMETPKILMKENPEIYERAG 162
Db 121 VWKLKMHNAQODKANAINEAKKEGFAFLPRYGKISSEMMIAKVMQILDEADVYNRD 180
QY 163 QFEDLADFLTKRATGDLARSVCYTCTKWTVAHNRMDPYFTTGLAELEDEFITIGH 222
Db 181 QFLEATDVIYSOMGTIKVNSCTAGYKAIWKKREGYPSNEFYALD--PRLEHLTTTKLRG 239
QY 223 HIVSPGTPCGNGILTAQAAAMGILLPGPRAVGLIDAHAG-----GIGTVEGEGALNMLAY 278
Db 240 DIYPLGSRAG-GLPREMAEKNGILPGIAVAVGANDDAIAPRAYGVTPG-----KLVM 291
QY 279 VFGTSSCTMASTTSPSFVPGVMPYYSAMVPGMLVVEGQSAAGALIDQLDFHRAV--- 335
Db 292 AMGSTISMLLGEDEQVEEGMCGVVEDGIIIPGYLGEGASAGVDIPAMVVKHGSATF 351
QY 336 EEARERAMQVNPPLRVWLAIRILEKTAQ--PSDAVALAKGLHYVRELGRRAPRADNHAR 393
Db 352 DEAOEKVNNVHALL-----EKASQLRPGES-----GLALDMWNGNSIILVDLELS 398
QY 394 AVICGLGMRDLDNLALYIAGLCIGYGLRQILDQAOTAGVSKNIVISGGAQO--HPLV 452
Db 399 GMLLTGYLQTKPRE--IYALALEATAFGTRALYDAHNGRGVNEHLXACGGRLQKKNHL 455
QY 453 RQILDATGCPVITTOCCPEVLLGSAIIGAVAGNIA---PSVGEAMQOF----- 498
Db 456 MQIFADVTNEIKVAAKQRPALGAAMFASVAAQSEVGYSDISEAAKMGWYKVDTEFKP 515
QY 499 --THVDKYYYPQERYOSLH-----RYRAYVQOLH 527
Db 516 IPEHVAIYERKLYQERYVTLHDYFGRGANDVMKRLKALKSIQH 556

RESULT 15
KIECRU
ribulokinase (EC 2.7.1.16) - Escherichia coli (strain K-12)
N:Alternate names: L-ribulokinase
C:Species: Escherichia coli
C:Date: 30-Jun-1998 #sequence-revision 31-Oct-1997 #text-change 01-Mar-2002
C:Accession: G64727; B29022; S40579; I41134
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.; Rose, D.J.; Mau, B.; Shao, Y.
S:Title: The complete genome sequence of Escherichia coli K-12.
S:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64727
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-566 <BLAT>
A:Cross-references: GB:AE000116; GB:U00096; NID:g1786240; PIDN:AAC73174.1; PID:g1786249;
A:Experimental source: strain K-12, substrain MG1655
R:Lee, N.; Gielow, W.; Martin, R.; Hamilton, E.; Fowler, A.
Gene 47, 231-244, 1986
A:Title: The organization of the arabid operon of Escherichia coli.
A:Reference number: A91559; MUID:87163495; PMID:3549454
A:Accession: B29022
A:Molecule type: DNA
A:Residues: 1-22, 'S', 24-126, 'RS', 129-349, 'S', 351-364, 'A', 366-402, 'S', 404-524, 'R', 526-566
A:Cross-references: GB:D10483; GB:J01597; GB:J01683; GB:J01706; GB:K01298; GB:K01990; GB
IDN:BA01334.1; PID:9216483
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A:Reference number: S40531
A:Accession: S40579
A:Molecule type: DNA
A:Residues: 1-22, 'S', 24-126, 'RS', 129-349, 'S', 351-364, 'A', 366-402, 'S', 404-524, 'R', 526-566
A:Cross-references: EMBL:D10483; NID:9216434; PIDN:BA01334.1; PID:9216483
R:Lee, N.; Carbon, J.
Proc. Natl. Acad. Sci. U.S.A. 74, 49-53, 1977
A:Title: Nucleotide sequence of the 5' end of arabid operon messenger RNA in Escherichia
A:Reference number: I41134; MUID:77102763; PMID:189315
A:Accession: I41134
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-14 <RES>
A:Cross-references: GB:K01304; NID:g145312; PIDN:AAA23465.1; PID:g145313
C:Comment: This enzyme catalyzes the phosphorylation of L-ribulose to L-ribulose-5-phosph
C:Genetics:
A:Gene: arab
A:Map position: 1 min
C:Superfamily: ribulokinase
C:Keywords: arabidose metabolism; phosphotransferase

Query Match 14.0%; Score 394.5; DB 1; Length 566;
Best Local Similarity 28.1%; Pred. No. 1.2e-21;
Matches 166; Conservative 75; Mismatches 221; Indels 129; Gaps 31;

QY 8 IGVNVSQSVRAGIFDLNGLSLSHAT-EKITTTRSGSRVDSQSEIWO--AVSCIRNA 64
DB 5 IGLDFGSDSVRALAVDC-----ATGEEIATNS-----VEMYPR--WQKGFCDAPRNNQ 49

QY 65 L-----TLADVCAQ--SVAGIGFATGSL-VVLDKNGDPLPYSPGCD 103
DB 50 FRHHRPDYIESMEALKTVLALSVGEORAAVVGIGVDSTGSTPAIDADGNVLAIRPEFA 109

QY 104 AKON--IIVMMDHRTBOERINATNHPVLN-----YVGGKISPEMERPKITLMLKENMPE 156
DB 110 ENPNAAPFLVMDHTVAEEETRLCHAPGVNDYSRITGIGYSSSEWPAKILHVTRODSA 169

QY 157 IYERAGOFFDLADELTTRATG----DLARSVCVTCKWTWLAHENRWDPDYFTIGL-- 209
DB 170 VAQSAASWIELCDWVPALSLSTTRPDIDIRGRCSAGHSLW--HES-WG-----GLPP 219

QY 210 AELADEDFIRIGHIVSP-----GTPCGNGLTAAQAAEMGLPGTPVAVGLIDAHAG 261

DB 220 ASFDELDPIINRHLPSLFDFTWTADIPVGT-LCPWAQRGLGPESVVISGAFDCHMG 278
QY 262 GIGTGVGEGALNNLAAYEGTSSCTMASTSPSF-----VPCWVGPPYSAMPGLMLVGG 317
DB 279 AVGA-----GAQPNMLVKYIGTSTCDILITADKOSGERAVKIGCGVDGSPVPGFGLG 334
QY 318 QSA-----AGAAIDLDLHPAVEAREMAQVNOPLPVWLADRLTEKTAQPSD 366
DB 335 OSAFGDIYAMFGRYLGMPLBQLAQHP-----ELKTIQINS-----QKOLLPAL 378

QY 367 AVALAKG-----LHVPR-FIGNRAPRADPHARAVICGLMERDNLIALYIGLCSTG 420
DB 379 TEAMAKNPISLDHLPVYLDWFMFGRRTPNANQRLKGYITDNLNLTADAPLFFGGLA 435

QY 421 YGLRQIIDLPQAGVVSKNITVSGG-AGQHPVRLQIADPCGIP---VITQOCCPEVLG 476
DB 436 FGARIMECFTDQGLAVNNVVALGSIARKNOVIMQACCDVLRPLQIVASDQC---ALG 492

QY 477 SAITGAVAGNIAPSVGEAMQOF-THVDKYVYP---QERYSLHRRYEAY 522
DB 493 AATFAAVAKVHADIPSAQOKMASAVEKTLQPCSEQAQRFQQL-YRRYQGW 542

Search completed: March 13, 2003, 16:56:01
Job time : 21.5662 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 10.6093 Seconds
(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208B-4
Perfect score: 2811
Sequence: 1 MTIKTVIGVGVSGSVRAG.....HHRREYAKYQLHTAKLRD 534

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448.5	16.0	560	1	ARAB_BACSU
2	447	15.9	564	1	ARAB_BACST
3	405.5	14.4	565	1	ARAB_ECO57
4	402	14.3	563	1	ARAB_BACDH
5	397	14.1	563	1	ARAB_MYCSM
6	394.5	14.0	565	1	ARAB_ECOLI
7	377	13.4	568	1	ARAB_SALTI
8	368.5	13.1	568	1	ARAB_SALTY
9	361	12.8	563	1	ARAB_YERPE
10	352.5	12.5	542	1	MP43_YEAST
11	340	12.1	545	1	ARAB_STAM
12	325	11.6	501	1	XYLB_LACPE
13	323.5	11.5	524	1	GLPK_MOUSE
14	323	11.5	530	1	YDEV_ECOLI
15	322.5	11.4	499	1	XYLB_BACSU
16	321.5	11.4	499	1	GLPK_SULSO
17	319.5	11.4	502	1	GLPK_PASMU
18	317.5	11.3	524	1	GLPK_RAT
19	312.5	11.1	505	1	GLPK_ENTCA
20	311.5	11.1	523	1	GLPK_HUMAN
21	310.5	11.0	554	1	GLPK_HUMAN
22	306	10.9	484	1	XYLB_ECOLI
23	306	10.9	498	1	GLPK_LACIA
24	304.5	10.8	494	1	GLPK_PRRAB
25	300	10.7	500	1	GLPK_ENTFA
26	297.5	10.6	502	1	GLPK_HAEIN
27	296	10.5	502	1	XYLB_LACBR
28	295.5	10.5	483	1	LYXK_PASMG
29	295	10.5	496	1	GLPK_BACSU
30	290.5	10.3	498	1	GLPK_ECOLI
31	290	10.3	495	1	GLPK_THERO
32	286	10.2	553	1	GLPK_HUMAN
33	283.5	10.1	513	1	GNMK_BACSU

34	282	10.0	503	1	GLPK_PSETO	087924 pseudomonas
35	279.5	9.9	497	1	GLPK_BACDH	09kdw8 bacillus ha
36	278.5	9.9	492	1	GLPK_AQUAE	066746 aquifex ae
37	276.5	9.8	497	1	GLPK_PYRRO	093623 pyrococcus
38	271.5	9.7	496	1	GLK2_THEMA	09x1ed thermotoga
39	271.5	9.7	501	1	GLPK_ECOLI	09x1ed escherichia
40	271.5	9.7	502	1	GLPK_CAEEL	021944 caenorhabdi
41	270.5	9.6	496	1	GLPK_THEAO	09wx23 thermus aqu
42	270.5	9.6	513	1	GNMK_BACLI	09x1ed bacillus li
43	267.5	9.5	508	1	GLPK_MYCLE	09c8b1 mycobacteri
44	263.5	9.4	497	1	GLPK_RHIME	086033 rhizobium m
45	261	9.3	501	1	XYLB_LACIA	09c1f8 lactococcus

ALIGNMENTS

RESULT 1
ARAB_BACSU STANDARD: PRT; 560 AA.
AC P94524: 005185;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (Ec 2.7.1.16).
GN ARAB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97237725; PubMed=9084180;
RA Sa-Nogueira I M G, Nogueira T V., Soares S., de Lencastre H.;
RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,
RT genetic organization and expression";
RT Microbiology 143:957-969(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;
RA Wipat A., Carter N., Bignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emerson P.T., Harwood C.R.;
RT "The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress
RT responses, the utilization of plant cell walls and primary
RT metabolism";
RT Microbiology 142:3067-3078(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Bortiss R., Boursier L., Brans A., Brun M., Bignell S.C., Bron S.,
RA Brouillett S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidis A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Ondega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengruber T.,
 RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [4]
 RP "TRANSCRIPTIONAL REGULATION."
 RX PubMed:10417639.
 RA Mota L.J., Tavares P., Sa-Nogueira I.M.G.;
 RT "Mode of action of Arar, the key regulator of L-arabinose metabolism
 RT in Bacillus subtilis.";
 RL Mol. Microbiol. 33:476-489(1999).
 CC "CATABOLIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate."
 CC "1- PATHWAY: L-arabinose catabolism; second step."
 CC "1- INDUCTION: Transcription is repressed by glucose and by the
 CC binding of arar to the operon promoter. L-arabinose acts as an
 CC inducer by inhibiting the binding of arar to the DNA, thus
 CC allowing expression of the gene."
 CC "1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY."
 CC "-----"
 CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)."
 CC "-----"
 DR EMBL: X89408: CAAG1586.1: -
 DR EMBL: 275208: CAAG9588.1: -
 DR EMBL: 299118: CAB14839.1: -
 DR Subtilase: Bg11905; arab.
 DR Pfam: PF00370; FGGY; 1.
 DR TIGRFAMs: TIGR01234; L-ribulokinase; 1.
 DR Transferrase: Kinase; Arabinose catabolism; Complete proteome.
 KM TRANSFERASE: Kinase; Arabinose catabolism; Complete proteome.
 FT CONFLICT 59 P -> F (IN REF. 1).
 SQ SEQUENCE 560 AA; 61009 MW; DE721AB73AF0AEF CRC64;

Query Match 16.0%; Score 448.5; DB 1; Length 560;
 Best Local Similarity 27.1%; Pred. No. 4,4e-25;
 Matches 153; Conservative 95; Mismatches 237; Indels 79; Gaps 20;

QY 8 IGVGVGSGVRAGIEDL-NGSLSHATEKT-----TTTRSGSRVE-----QSSQELIWQ 55
 DB 5 IGVDEFTLSGRAVLVHQTGEELAAVKEYNHAAVIDVLPRTGOKLPDMALQHPADYLE 64
 QY 56 AVCSIRNALTLADVCQSVAGIGFDPAT-CSLVVLDKNQDPLPVSPEGDAKON--IIYVM 112
 DB 65 VLETTTISLEQGVGVDRKDIIGIDITACTITLIDSSGQPLCLPXEERPHSYVLLMK 124
 QY 113 DHRATEAERIN-----ATHRPVLNVYVGKISPEMETPKILMLKENNPEIYERAGQFDLA 168
 DB 125 HHAQKRIADRLNQIAEEGEAFLORYGKISSSEMIKPKVQIAEAENHIYEAADRITIEA 184
 QY 169 DFLTWRTATGLANSVCVTCTCKWTYLAHNNRMDPYFTTIGLA-ELADEDFRIGHNIVSP 227
 DB 185 DWIYYQGLSGSLKRSNCTAGYKAMWSEKAGYPSDQFEKLPNSMTITKD--KLSGSIHSV 242
 QY 228 GTPGNCILTLQAAAEMLLGTPTPAVGLIDAHAGCIGTGV-EGGALNNLNAVYGTSSCT 286
 DB 243 GKRAGS-LTEKMAKLLIGLPGTAIVANVDAHV-SYPAVGTTEPG--KMLMTIGSTCH 297
 QY 287 MASTTSPFVGVGWPYYSAVPGMLVLEGOSAAAGAIQQLD--PHPAV-EEARE--- 340
 DB 298 VLLGEEVHIYRGMGVVNDNGLPGYAGYEAQSCVGDHPFMFYKTCVPRAVYQEAKEKNI 357

QY 341 -----MAQRYNOPLPWLADRIIEKTAQPSDAVALAKGLHVPEFLGNRAPFADPHARAV 395
 DB 358 GVHELLSEKKNHQAQPG-----ESGLLALDMWNNRSTLVADADLTGM 398
 QY 396 ICGIGMERDLDNLALYIAGICGIGYGLROILOAQTVQGVSKNITYISG-ACQHPLVQ 454
 DB 399 LIGWTL--LTKPEIRALVEATVYCTRMIIETFKESGVIEELFAAGIAEKNPVMQ 455
 QY 455 ILADTCIPVITTCOCPPVLGSAIIGAVG-----NAPSVGEAMQGFTHVDK 503
 DB 456 IYADVTMMDIKISGSPAPALGSAIFGALAGKEKGYDDIKKAAANMGLK-----DI 509
 QY 504 YVYPOERYGSLHRRYEAQKOLH 527
 DB 510 TYTPNAENAAVYEXLVAYEKLVIH 533

RESULT 2
 ARAB_BACST
 ID ARAB_BACST STANDARD; PRT; 564 AA.
 AC Q9S468;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 OS ARAB.
 GN Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SYRAIN-T-6;
 RA Gillead-Gropper S., Shoham Y.;
 RT "The L-arabinose utilization gene cluster from Bacillus
 RT stearothermophilus T-6.";
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC "1- CATABOLIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate."
 CC "1- PATHWAY: L-arabinose catabolism; second step."
 CC "1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY."
 CC "-----"
 CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)."
 CC "-----"
 DR EMBL: AF160811: MAD4517.1: -
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY; 1.
 DR Pfam: PF02782; FGGY; 1.
 DR TIGRFAMs: TIGR01234; L-ribulokinase; 1.
 KM TRANSFERASE: Kinase; Arabinose catabolism.
 SQ SEQUENCE 564 AA; 62202 MW; DBFPDFBC36EDDB5 CRC64;

Query Match 15.9%; Score 447; DB 1; Length 564;
 Best Local Similarity 28.5%; Pred. No. 5.7e-25;
 Matches 159; Conservative 89; Mismatches 245; Indels 64; Gaps 20;

QY 5 KTVIGVDVSGSVRAGIEDLNGS-LLSHAT-----EKITTRSGSRVE-----OSSO 51
 DB 4 KYVIGIDYGTESGRAVLVDELEGNIDAHVYRPHGVIDEVLP--ESNVQLEPDMALQHPA 61
 QY 52 EIMQAVSCIRNALTLADVCQSVAGIGFDPAT-CSLVVLDKNQDPLPVSPEGDAKON--I 108
 DB 62 DYIEVLATAVPTVQLQSGVPADVLTGIDTSCVTMLPVASGEPLCLKPEFRNRRPSWV 121
 QY 109 IYVMDHRTAQDAERIN-----ATHRPVLNVYVGKISPEMETPKILMLKENNPEIYERAGQ 164
 DB 122 KLMKHNAAQDEANILNEIAKRGAEFLPRYGGKISSSEMIKPKVQIAEAENHIYEAADRIT 181

[illegible]

RESULT 3			
ARBAB_ECO57	STANDARD:	PRT:	565 AA.
ID	ARBAB_ECO57		
AC	PS6541:		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	L-ribulokinase (EC 2.7.1.16).		
GN	ARBAB OR 20072 OR ECS0067.		
OS	Escherichia coli O157:H7.		
OC	Bacteriia; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
ON	NCBI_TaxID=83334;		
XX	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=O157:H7 / EDL933 / ATCC 700927;		
RC	MEDLINE=21074935; PubMed=11206551;		
RC	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,		
RA	Postif G., Hackett J., Klink S., Boulin A., Shao Y., Miller L.,		
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,		
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,		
RA	Weich R.A., Blattner F.R.;		
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";		
RL	Nature 409:529-533(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=O157:H7 / RIMD 0509952;		
RC	MEDLINE=21156231; PubMed=11258796;		
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,		
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,		
RA	Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,		
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;		
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli		
RT	O157:H7 and genomic comparison with a laboratory strain K-12.";		
RL	RNA. Res. 8:11-22(2001).		
CC	-1. CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-		
CC	phosphate.		
CC	-1. PATHWAY: L-arabinose catabolism; second step.		
CC	-1. SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/announce/
CC or send an email to%5Blicense%5B-sib.ch%5D](http://www.isb-sib.ch/announce/or_send_an_email_to%5Blicense%5B-sib.ch%5D)).

CC -----

DR EMBL: AE005183; AAC54367.1; -
DR EMBL: AP002550; BAB33490.1; -
DR TIGRFAMs: TIGR01234; L-trioloXinase; 1.
KM Transferrase; Kinase; Arabinose catabolism; Complete proteome.
FT
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 565 AA: 61096 MW: 45E7D9AA096B9CAD CRC64:

Query Match	14.4%	Score 405.5	DB 1	Length 565
Best Local Similarity	27.7%	Pred. No. 5.8e-22		
Matches 163; Conservative	77	Mismatches 224	Indels 125	Gaps 29

QY	8	IGVDGSGSVRGTGFDLNGSLSHMTKTTTRRSGSRVEDSSOGLTQ--ANCSSIRNAL	65
Db	4	IGLDGSDSVRALAD-----CTTGEETATSYEMVPR--WVGQGFCDAPNNQF	49
QY	66	-----TLADYCAQ--SVAGIGFDATCSL-VVLDDKNGDPLPVSPEGRA	104
Db	50	RHHPRDYIESMEALKTYTALIELSVGBRAAVUGIGVDSTGSPARIDAGGNVLAIRPEAF	109
QY	105	KON--TIYMMHRAEQAEIRINATHHRYLN---YVGCKISPEMETPKILMKEMPEI	157
Db	110	NNENAFVLMKDHATAEAEAEELTRILCHAPGNVDYSRYIGISYSEFMFAKILHVTRODSAV	169
QY	158	YERAGQFDLAFELFWRATG-----DLARVCTVCTCKWTLAHENRMKDPDYERTIGL--A	210
Db	170	AQSAASWIELCDWVALLSGTTRPQDIRRGCSAGAKSLIM--HES-WG-----GLPPA	219
QY	211	ELADEDFIRIGHIIVSP-----GTPGNGILTAQAAEMGLRPTRYAVGLIDAHAG	262
Db	220	SFDELDPLRNHLRSLPFTDTWTADIPVGT--LCPDMAQRILAPRESVVISGAFDCHMGA	278
QY	263	IGTVGVEGALNNLNAVVGTSCTMASTTSPF-----YRGVNGPRYYSAAVPRGLMVEGQ	318
Db	279	VGA---GAQPNALVKVIGTSTCDLLADKQSVGERAKVKGIGQVDGVSVPFGISLEAQ	334
QY	319	SA-----AQAIDLDLHFHRAVEAEAREMAQVRQPLP-----VLADRIELEKTAO	363
Db	335	SAFGDIYAFMGFNVLRMPLEQLOAGHRELKEQIDAQOK--QLLRALTEMAKRPISLDH---	389
QY	364	PSDAVALAKGLHVRDE-FLGNRAPRRADHRAVATGCLGEMEROLDMLLYTAGLGGITG	422
Db	390	-----LEVVLDWFGMRTRPRANORLKKVITDILNLTAPARLLFGGLD--ATAFG	436
QY	423	LKQILDAQTAQGVNASKNIYISGG-AGQNRHLYMQLIADPGSP--VITTPQCEPVLLGSA	478
Db	437	ARAEECTTDSGIAVNNVVALGSIARKKNOVLIQACCDVLNRPLOIYASQCC---ALGAA	493
QY	479	ILGAVAGNIAPSVSGAMQOF-THVDKYYUYQ---ERYOSLIHRRYEAV	522
Db	494	IPAAVAAKVHADIPSAQOKMSAVAEKTIQPRSEQAKQREQL--YRRYQW	541
RESULT 4			
ARAB_BACHD	ARAB_BACHD	STANDARD:	PRT: 563 AA.
AC	Q9K8Q3:		
DT	15-JUN-2002 (Rel. 41, created)		
DT	15-JUN-2002 (Rel. 41, last sequence update)		
DT	15-JUN-2002 (Rel. 41, last annotation update)		
DE	L-ribulokinase (EC 2.7.1.16).		
GN	ARAB OR BH1872.		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxId=86665;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C-125 / JCM 9153;		
RX	MEDLINE=20512582; PubMed=11058132;		

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001513; BAB05591.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR Pfam: PF02782; FGGY_C; 1.
DR TIGRfams: TIGR01234; L-ribulokinase; 1.
KW Transferase; Kinase; Arabinose catabolism; Complete proteome.
SQ SEQUENCE 563 AA; 61668 MW; CCBC39006669ACD CRC64;

Query Match 14.3%; Score 402; DB 1; Length 563;
Best Local Similarity 26.7%; Pred. No. 1e-21;
Matches 155; Conservative 82; Mismatches 260; Indels 84; Gaps 20;

QY 4 TKTVIGVDSGSVRAGIFDL-NG-SLSSTATE-----KITTRRSGSVESSQ 51
DB 3 TKTVIGVDSGSVRAGIFDL-NG-SLSSTATE-----KITTRRSGSVESSQ 51
QY 52 EIMQAVSCIRNALTLADVCASVAGIGFDP-CSLVLDKNGDPLRPSPEGDAKON-- 107
DB 63 DYVEVLITSVPAVMKESGVADVDYIGIGVDTACTMLRYPDESGPCLLQA--YKDNPHS 120
QY 108 -IIVMDHRAATEQAEERIN---ATNHPVLTNVGKISPEMETPKILMKENMPETIERAG 162
DB 121 WVKLMKHNHAADKANAINEMAEKRGFAFLPRYGKISSEMMIAKVMQDILDEADVDVNRTD 180
QY 163 QFEDLADFLTRARAGDLARSCVTCTCKWTMLAHENRMDPROFRTIGLAELADEPFIIGH 222
DB 181 QFLEATWIVISQMTGKTVKKNCTAGYKAIMHRRGYPSEFEKALD-PRLEHLLTTTRLRG 239
QY 223 HIVSPGTRPCNGTLTAQAAAEKGLRPTPVAVGLIDAAG---GIGTVYEGGALNNLAY 278
DB 240 DIYVLGERAG-GILRPEAEKGLNPGIAVAVGNDAHAAPVAVTTPG-----KLYM 291
QY 279 VFGTSSCTMASTTSPSFVPGVWGPYSAWPGMLVEGGOSAGAALDQLDFHPAY--- 335
DB 292 AMGTSICHHMLLGEKEQEVGECGVEDGIIIPGYLGVEAGSAGVDIPAMFVKHVSATF 351
QY 336 EEAERMAQRNQLRPVLAIRILEKTAQ--PSDAVALAKGLHVPPELGNRAPADPHAR 393
DB 352 DEAEQKGVNVAHLE-----EKASQLRGES-----GLLADDMWNGNSIIYDTEL 398
QY 394 AVICGLMERDLDNLALYTAGLGIGYGLROIIDAOTAGVSKNIVISGAGQ-HPV 452
DB 399 GMLGTYLQTKRPE---TYRALLEATFAGTRAVDAFHGRQVEVHELYACGGRLPKNNHLL 455
QY 453 ROIADLTGIPVITTOCEPVLGSAIIGAVAGNIA---PSVEAMQOF----- 498
DB 456 MQJADVTNEIKVAASKQRPALGAAMFASVAASEVGVGDSIDEAAKKMGKRVADETFKR 515
QY 499 --THVDKTYTPQERTYSLH-----RRPEAYKOLH 527
DB 516 IPEHVAITEKLYOEYVTLHDYFGRGANDVMKRLKALKSIOH 556

RESULT 5

ARAB_MYCSM
ID ARAB_MYCSM STANDARD; PRT; 563 AA.
AC Q9LBO3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB.
OS Mycobacterium smegmatis.
OC Bacteria: Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1772;
RN [1]
RC SEQUENCE FROM N.A.
PC STRAIN-SMDU;
RA Soda A., Takata G., Izumori K.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB038527; BAB92524.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR Pfam: PF02782; FGGY_C; 1.
DR TIGRfams: TIGR01234; L-ribulokinase; 1.
KW Transferase; Kinase; Arabinose catabolism.
SQ SEQUENCE 563 AA; 60432 MW; 12F097597DD07918C CRC64;

Query Match 14.1%; Score 397; DB 1; Length 563;
Best Local Similarity 29.3%; Pred. No. 2.4e-21;
Matches 166; Conservative 64; Mismatches 249; Indels 88; Gaps 19;

QY 5 KTVIGVDSGSVRAGIFDL-NGSLSHATEKITTRRSRVE----- 47
DB 8 KYVIGVDFGLSGRALVRAVDGAEKSAER---TYRHGVTEALPGRPEVRLPADYAL 63
QY 48 QSSQEIQAQVSCIRNALTLADVCASVAGIGFDP-CSLVLDKNGDPL-----PVSP 100
DB 64 QVRADYIDLRFALRPAVAVNAGIDPADVIGIGTDFCTACTVAAATSDGTPICQDEFAADRP 123
QY 101 EGDAKNIIVMDHRAATEQAEERIN---ATNHPVLTNVGKISPEMETPKILMKENMP 136
DB 124 HAYAK---LWRNHAAPQADRIINALAARGETWLPYGGILISEWFAKGLDLEEDPE 179
QY 157 IYERAGQFPLADFLTRARAGDLARSCVTCTCKWTMLAHENRMDPROFRTIGLAELAD-- 214
DB 180 IYAADRWVGCADWIVQTLGRYVRNISTAGYK-----AIRDQKYPSPAFLAELNPGF 223
QY 215 ----EDFIR--IGNHIVSPGNGTLTAQAAAEKGLRPTPVAVGLIDAHAGIGTVGV 268
DB 234 ASFVEDKITEQPIGLGSAAGT-----LTAQAAATGTPREGIIVAVGNIDAHVYRAADAA 288
QY 269 EGALNNLAVVFGTSSCTMASTTSPSFVPGVWGPYSAWPGMLVEGGOSAGAALDQL 328
DB 289 DPG---RLAIIMGTSTCHVNMGRFLREVPCGVVDGIGTIDGLGVBAQSGVDIPAMF 345
QY 329 LDFHPAVEEREMAQVRNQLRPVLAIRILEKTAQ--PSDAVALAKGLHVPPELGNRAP 388
DB 346 TKNCVPKEIATEASRR-----GLTMEHLSLAEDQEVG---EHLGVALDMHSGNSRVLV 397
QY 389 DPHARAVICGEMERDLDNLALYTAGLGIGYGLROIIDAOTAGVSKNIVISGAGQ 448
DB 398 DHHLSGIMVQOTLD--TTCVDDYTRALLEATFAGTRAVETFORSGVPEVELVYAGGLIK 454

ID ARAB_SALTY STANDARD: PRT: 568 AA.
 AC P58542;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR STY0120.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18.
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felkell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagers K.,
 RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL627265; CAD01260.1;
 DR TIGRFS: TIGR01234; L-ribulokinase; 1.
 KM Transferase; Kinase; Arabinose catabolism; Complete proteome.
 FT INIT_MET 0
 SO SEQUENCE 568 AA; 61643 MW; A1515DB7249A0F75 CRC64;
 Query Match 13.4%; Score 377; DB 1; Length 568;
 Best Local Similarity 26.4%; Pred. No. 6,8e-20;
 Matches 151; Conservative 82; Mismatches 249; Indels 90; Gaps 24;
 QY 8 IGVGVGSGYRAGTFD-LNGLSLSHATEKITTRRS-----GSRVQSSOEIWOAVCSC 60
 DB 4 IGVGVGSDYRALVADCATDEIATSVEMYLRMQEGRYCDGPNQFHHHPRDYMESMAA 63
 QY 61 IRNALT-LADVCAQSVAGIGEDATCSL-VVLKNGDPLVPSPEGDARON--IIVMMDRA 116
 DB 64 LKAVLAQLSAQRANVYGVIGVSTGPAPIDAGNVLAPEAPENPMAYLMDKHTA 123
 QY 117 TEOAERINATNH-----PVLNVYGGKISPEMERPKILMLKENPEIYERAGOFEDLDFL 171
 DB 124 VEEADETRLCRHPKGVKDYRYIGIYSSSEMFMAKILHYRQDSAVNAQAVSWELCDWV 183
 QY 172 TWRTATG-----DLARVCYTYCKWTWLAHENRMDPDYFRITGL--AEIADDFIRICHHI 224
 DB 184 PALISGTRPDIDIRGRCSAGHKTLM--HES-WG-----GLPRPASFDLPCINRHL 233
 QY 225 VSP-----GTPCGNGTLAAQAAEMWLLPGTPVAVGLIDAHAGIGTYGVEGALNNL 276
 DB 234 RYPLFSTFTPADLPVGL-LCAEMHQRIGLDESVYISGAFDCHHGAAGA-----GAQPNLT 288
 QY 277 AVYFGTSSCTMASTTSPSF-----VPGVWGPPYSAMVPGMLLVESGQSAAGA----- 323
 DB 289 VKVIGSTCOILLADKQSVGDRAVKGICGQVDSVYVNFILGLEAGQSAFDIYAMFSRVL 348

QY 324 --AIDOLDHPRAVEEREMAKORVNOPLVWADRIIEKTAQPSDVAALAKGLHWPE-- 379
 DB 349 SWPLEQIAQHPBELKTJINSQK--QLPLV-----TDAMAKRPSLDHLEVLV 394
 QY 380 --FLGNAPAPADPARAVITGIGMERDNLALYIAGICGIGLQIIDAOAGVGS 437
 DB 395 DWFGRRTPNANORLKGVIITDLNLTADP--ALFGGLVSTAFGARAIQCEFTQGIAY 451
 QY 438 KNIVISGG-AGQHPVLVQIADTCGIP--VITQCCPEVLLGSAILGAVAGNIAPSGE 493
 DB 452 NNWVALGIAKKNQVIMQVCCDVLNRPLOIYASQCC--ALGAAITPAVAAYAKVHADIPA 508
 QY 494 AMQGF-THVDKYRYPQERTQSLHRRREAYKQ 524
 DB 509 AQSMAASAVERTLRPREQARFERLYRRYQO 540
 RESULT 8
 ID ARAB_SALTY STANDARD: PRT: 568 AA.
 AC P06188;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR STW0103.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85232044; PubMed=2989100;
 RA Lin H.-C., Lei S.-P., Wilcox G.;
 RA "The arabid operon of Salmonella typhimurium LT2. I. Nucleotide
 RT sequence of arab and primary structure of its product,
 RT ribulokinase";
 RL Gene 34:111-122(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M11047; AAA27023.1;
 DR EMBL: AE008698; AAL19067.1;
 DR PIR: A24984; A24984.
 DR ScyGene: SG10013; arab.
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR Pfam: PF02782; FGGY_C_1.
 DR TIGRFS: TIGR01234; L-ribulokinase; 1.
 KM Transferase; Kinase; Arabinose catabolism; Complete proteome.
 FT INIT_MET 0
 SO SEQUENCE 568 AA; 61643 MW; A1515DB7249A0F75 CRC64;

SQ SEQUENCE 568 AA; 61620 MW; B1D838C37B7C4134 CRC64;
 Query Match 13.1%; Score 368.5; DB 1; Length 568;
 Best Local Similarity 26.8%; Pred. No. 2.8e-19;
 Matches 156; Conservative 79; Mismatches 235; Indels 113; Gaps 27;
 8 IGVNDSGSGVRAGIDPLNGSLSHAT-EKITTTRSGSHVEOSS-----50
 4 IGLDGSQSVRLAYDC-----ATGDEIATSVEMYPWMQSGRCYCDGNQFRHHPRDY 56
 51 QEIMQAVSCINNLTLADYCAOSVAGIFDPTCSL-VLIDKNDPLVPSPEGDAKON--107
 57 MESMAKALAVLAOLSAQA--RANVVGIGVSTGSTPAIDAGNVLALRPEFAENPM 114
 108 IIVMMDHRTAEQERINAHN-----PVLNVGKRTSPEMETPKTLMKENPEIYERAG 162
 115 FVLMDHRTAEQERINAHN-----PVLNVGKRTSPEMETPKTLMKENPEIYERAG 174
 163 GFEDLADFLTWRTG-----DLARSVCTVCKMTWLAHENRMDPDYFRTIGL--AELADE 215
 175 SWIELCDWVPALLSGTTRPQDIRGRCSAGHKTW--HES-WG-----GLPRAFPDE 224
 216 DFRIGHNIVSP-----GTPCCNGLTQAQAAEMGLLPFPVAVGLIDAHNGIGTVG 267
 225 IDPCINRHLRYPLFSETFTADLPVGT-LCAEWAQRIIDLEPSYVIGSGAFDCHGAVGA--281
 268 VEGALNNLAAYVGTSSCTMASTTSPSF-----VPGWVGYSYAMVSGMLVVEGQSAAGA 323
 282 --GAOPNPLVKYIGTSTCILADKOSVGDRAVKIGGVGDSVWPNFTGLEGQSAFED 339
 324 -----AIDOLDHPAVEAREMAQVNOPLPVWLADRLILEKTAQPSDAVALAK 372
 340 IYAMFSRVLSWPLEQALQAHPELKQINASK--QLLPAL-----TDMAKNP 385
 373 GLHYVPE-----FLGNRAPADPHARAVICGLMERDNLALYTAGLGIGYGLROIID 428
 386 SLIDHLPVLDWENGRRTPANORLKGVIIDMLATDAP--ALFGLVASTAGARAIIE 442
 429 AOTAGVSNKNIVISGG-AGOHPLVROILLADTCGIP--VITTOCCPEVLGSAIIGAVA 484
 443 CTDDGIANNNMALGCIARKKQVIMQCCDVILNRPLOIVASDOCC--ALGAATTAAYA 499
 485 GNIAISVGEAMQOF-THVDKYYYPQ---ERYQSLHRRYEAV 522
 500 AKVHADIPAAQOSMASAVERTLRPHEDQARFEQL-YRRYQOW 541
 RESULT 9
 ARAB_YERPE STANDARD; PRT; 563 AA.
 ID ARAB_YERPE STANDARD; PRT; 563 AA.
 AC P58543;
 DT 15-JUN-2002 (Rel. 41, Last Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR YPO2254.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 OX NCBI_TaxID=632;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Blovat Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebalnia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.",
 RL Nature 413:523-527(2001).

CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AJ14151; CAC91058.1; ALP INT.
 DR TIGR: TIGR01234; L-ribulokinase; 1.
 KW Transferase; Kinase; Arabinose catabolism; Complete proteome.
 SQ SEQUENCE 563 AA; 61833 MW; 88FE1EB00A5414E CRC64;
 Query Match 12.8%; Score 361; DB 1; Length 563;
 Best Local Similarity 26.4%; Pred. No. 9.7e-19;
 Matches 152; Conservative 81; Mismatches 257; Indels 86; Gaps 25;
 8 IGVNDSGSGVRAGIDPLNGSLSHATEKITTT-----RRSGRVEQSSOEIWO 55
 10 IGLDFGSDSVRLAYD-----CONHTEIDTEVYVPRMKKGLYCAQAOQFRHHPLDYIE 64
 56 AVSCICRNL-PLADYCAOSVAGIFDPTCSL-VLIDKNDPLRVPSEDAKON--IIVW 111
 65 ASEQAIRQWKRILSEQRORHIVGIGVSTGTPAPRIIDEGQVIALRPDRADNPAMFVLM 124
 112 MDHRTAEQERIN-----ATHHPVLNVGKRTSPEMETPKTLMKENPEIYERAGQFPD 166
 125 KQHTALEEAEINRLORSEFADYSYIGVYSSEFMFKILHVTADAVRAAASWIE 184
 167 LADFLTWRTG-----DLARSVCTVCKMTWLAHENRMDPDYFRTIGLAELEDEFTIRG 221
 185 LCDWVPALLSGTTPADODIRGRCSA-----GHKSLMPWSWGLPRAFLALDTSLVN 237
 222 H-----HIVSGTCCNGLTQAQAAEMGLLPSTVAVAGLIDAHNGIGTVGEGALN 274
 238 DLDYRPHFTPTTYTAEVQO-ITAEWAERLGLPTVYLISGAFCHMGAVGA--GAOPY 292
 275 NLAYVEGTSSCTMA-----STTSPSFVPGWGPYASAVPQIWL-VEGQSAAG--AAID 326
 293 TLVKYIGTSTCIDLIADDQRYGDRALTAGICQYEGSVLG-WIGMEGQSAFEDYAMFS 351
 327 QILDF--HPAVEARMAQVNOPLPVWLADRLILEKTAOPS-DAVALAGLHYVPEFLN 383
 352 NLSWPLHQAALTOPWQOLKQIESNLSLTRAQONPSIDHLPV-----VLDFWNGR 406
 384 RAFPADPHARAVICGLMERDNLALYTAGLGIGYGLROIIDQAGVSNKNIVIS 443
 407 RFPNANORLKGVIIDMLATDAPFLGFLA--ATAFGARALMBECPDODIPIDNVLAL 463
 444 GG-AGOHPLVROILLADTCGIP--VITTOCCPEVLGSAIIGAVANINAPVGEAMQOFT 499
 464 GGIARSPYIMQYCADVMNRPLDIVASDOCC--ALGAATFAVAGAHNDVPTAQRHMA 520
 500 -HVDKYYYPQ---RYQSLHRRYEAYKOLHTAK 530
 521 CNIEFTLIPDPVQVYRYQOL---YORYQOWCHTAE 552
 RESULT 10
 MP43_YEAST STANDARD; PRT; 542 AA.
 ID MP43_YEAST STANDARD; PRT; 542 AA.
 AC P53583;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MP43 protein.
 GN MP43 OR YML249C OR N0875.
 OS Saccharomyces cerevisiae (Baker's yeast).


```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5286C / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeher U., Behnhauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RN Yeast 13:849-860(1997).
RN [2]
RP SEQUENCE OF 41-542 FROM N.A.
RC STRAIN=ENY.WA-1A;
RA Boles E., Hettmann C., Zimmermann F.K.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X96722; CA65495.1; -
DR EMBL: 271525; CA96156.1; -
DR EMBL: X94214; CA63905.1; -
DR SGD: S0005193; MPA43.
FT CONFLICT 46
SQ SEQUENCE 542 AA; 61667 MW; 830A8B4E95365AD CRC64;

Query Match 12.5%; Score 352.5; DB 1; Length 542;
Best Local Similarity 25.9%; Pred. No. 3.8e-18;
Matches 127; Conservative 78; Mismatches 190; Indels 95; Gaps 22;

QY 8 IGVGVGSGSRAGIFD-LNGLSLSHATEKITTRRSQS---RVEQSOEIMQAVSCIRN 63
DB 9 IGVGVGSSARIGVYNYNDALLEMAQEPVYODSSKSKMKFQKSTETIKALOKCLOK 68
QY 64 ALTLADYCAQSVAGIGFADATCSLVVL--DKNGDPLVPSPEGDAKONITIVMDHRTAQAE 121
DB 69 -----LNIREVEKSCGVSATCSLAIFERDRTSNMLIPYND--NATFWDSSAVNECQ 121
QY 122 RIN-ATHHPVNLVYVGKISPEMETPKILMKENKPEIYERAGOFFDLADEFLTWRTGDLA 180
DB 122 WLNMQCPQQLLDYLGKFEVPEMGVPLKFLDEYSHLRDKHFHFDLHYIAY---ELS 177
QY 181 RSVCTVYCKTW-----LAHEN-----RMDPDYFRTIGLAELADEDFRIGHNI 224
DB 178 R-----LEWNIIEGLLGRENLNGIGNDGEVSGWSSSFYKNI-----INLPSN- 219
QY 225 VSPGTCGNGLTQAQAAEMGLPCTPVAAGLIDAHAG--GIGTVGVEGALNNLNAVYEGT 282
DB 220 VSIQGT---TSLVANKH-----STYVRSCLDSYASMFVAVASPHLE---TSLFMAGT 266
QY 283 SSCTMASTT-SPEFVPGWGPY-----YSAMVPGMLVLEGQSAAGAIDOLL 329
DB 267 SSCMWTCTTSDIRIPGWPGRPDTIIDNRGDFSYAA-----GQSGTGLIEHLEF 316
QY 330 DHPRAVEAREMAQRVQPLPWLADRILEKTAQPSDAVALAKGLHVPFLGRAPAD 389
DB 317 ESHPCARKIILKDGADIVQVLEQFTRD--IEKNNGLSITH-LTKDMFEYGYEGNRPFPAD 373
QY 390 PHARAVICIGMERDLNLLATYIAGCGIGYGRQILDA--QTAQVAVSKNVIYSAGG 447
DB 374 PRIGSGTGSTDSMLNLYKYICILEFSLFQTKLIIDTFQENSNHIIKEKISGSA 433
QY 448 QHPLVROILA 457
DB 434 KNERLJSLIS 443

```

```

RESULT 11
ID ARAB_STAM STANDARD; PRT; 545 AA.
AC 099W57;
DT 15-JUN-2002 (Rel. 41, Last Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB OR SAV0552 OR SA0510.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cul L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT aureus.
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RL Lancel 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003359; BAB56714.1; -
DR EMBL: AP003130; BAB41741.1; -
DR InterPro: IPR000577; FGGY_Kin.
DR Pfam: PF00370; FGGY_1.
DR Transferrase; Kinase; Arabidose catabolism; Complete proteome.
SQ SEQUENCE 545 AA; 60965 MW; E3917036237CEA4B CRC64;

Query Match 12.1%; Score 340; DB 1; Length 545;
Best Local Similarity 24.7%; Pred. No. 3.1e-17;
Matches 140; Conservative 109; Mismatches 245; Indels 72; Gaps 23;

QY 8 IGVGVGSGSRAGIFD-LNGLSLSH-----ATEKITTRRSGRSRYE-----OSSOEIMQAV 57
DB 5 IGVGVGASGRVFLINTNGQVYSKPYKPTTHGYISELNGKIPHYALQNSNDYIEIM 64
QY 58 CSCIRNALTLADYCAQSVAGIGFADATCSLVVL--DKNGDPLVPSPEGDAKQ-----NII 109
DB 65 EGISYIVRSKSIDPVNIVIGIDFTSSTIIFDEN-----LNVVNLKQFKNNPHAYVK 119
QY 110 VMMDHRTAEQERINAT---HHPVLNVYVGKISPEMETPKILMKENKPEIYERAGOFF 165
DB 120 LMKHHGAYKAEKLYQTAIENNNKWLCHYGYNVSSSEMMIKRIMVNNRAPEIMEKTYAIM 179
QY 166 DLADEFLTWRTAGDLARVCTVCTKWTWLAHENRMDPDYFRTIG--LAELADEDFRIGHN 223
DB 180 EAGQWIVNKLTKNNKVRNCGGFAFV--EETGTFHYDLFPKIDPKSKVYID--KVSAP 235
QY 224 IVSPGTCGNGLTQAQAAEMGLPCTPVAAGLIDAHAGIGTVGVEGALNNLNAVYEGTS 283
DB 236 VVNIQGEVYVK-LDDKMAQKQIKSTKEMTVSPFIIDAHASLLG-IGSEKD--KEMTMVAGTS 291

```



```

Db      1  MSAAVLGIDIGTSAVAVKSAIDKDGNNVVAQAASAKYALADQPHGYSQDEDEWVOTQAIR 60
QY      63  NALLIADYCAQSAVAGIGCFDATC -SLIVYLDKNGDPLPVSPEGDAKONITVMDHRTQDAE 121
      61  ELLQOGEVTAQDIEGLSISYSGMGHGLVLLDESAATV -RPAIIMNDTRTSSOCR 111
Db      122  RINATH -HPVINYGVGKISPEMERPKILMEKENPEIERAGOFDLDPLFWRATGOLA 180
      112  ELESQFGDDFKITIGNRPLEGFTLPKLLVAKENEPNIKRARTFLPDPDYLRRTGKLA 171
QY      181  --RSVCT-----VTCMWTLAHENRMDPDYFRTIGLAELADEDFIRIGHIYSPGTPC 231
      172  MDRSDATGCTVLDDITTSQWSEFLCNQDIP-----LFLCPPLISTAYVGH----- 217
Db      232  GNGLTQAQAAMGGLPCTPPVAVGLIDAHAGIGTGVGAGLNNIAY -FGTSSCTMA -S 289
      218  ---INQTYAQLSGSVNTRKVFEGADNAGAVGA---GILSDKALVSGTGVLLKYE 270
QY      290  TTSPSPFVGWVGPYSAAVPGLMLVEGQSAAGAIDDL-----DFHPAVEAREMA 342
      271  DNAQIDYRGVL-QYERIAFPCKYISMGTLLAGTSLNMFKOTFPAPDEDFGVVASAEQ-- 327
Db      343  QRVAQPLPFWMLADRIELEKTAQPSDAVALAKGLHVPERFLGNRAPADPHARAVICGL -GM 401
      328  -----STIGANGLLFPAYIVGEAPAPADATIRGSPFGVDS 363
QY      402  ERDDNLALTYIAGLCIGYGLRQLLDAAQTAAQGVASKNIVISGAGQHPLYRQLADTCG 461
      364  HQRADFVRAV---LEGIIFFSEFDILIKYOHNGAEFKTIVSIGGAKSALMLQIDAIJFN 419
QY      462  IPIVTTQCCSEVVLGGSALILGAVAGNIANSVGEAMQOFHVQKYYRPOERYOSLHRRYA 521
      420  CKVYSLKNEGPGMGAMATTAATGAGWEKFTLADCAQTFVHGKATPYTAHAQYQEMTRL 479
Db      522  YKOL 525
      480  YQOI 483

RESULT 13
GLPK_MOUSE
ID      GLPK_MOUSE      STANDARD;      PRT;      524 AA.
AC      064516;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      GlyceroL kinase (EC 2.7.1.30) (Atp:glycerol 3-phosphotransferase)
DE      (glycerol kinase) (GK).
CN      GK OR GYK.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BAIB/C;
RX      MEDLINE=97038697; Pubmed=8884278;
RA      Hug A.H., Lovell R.S., Sampson M.J., Decker W.K., Dinulos M.B.,
RA      disteche C.M., Craigen W.J.;
RT      "Isolation, mapping, and functional expression of the mouse X
RT      chromosome glyceroL kinase gene.";
RL      Genomics 36:530-534(1996).
CC      -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
CC      METABOLISM (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.
CC      -1- PATHWAY: glycerol utilization, rate-limiting step.
CC      -1- SUBCELLULAR LOCATION: BOUND TO THE MITOCHONDRIAL SURFACE OR
CC      CYTOPLASMIC (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE FPCOKINASE / GLUCOKININASE /
CC      GLYCEROKINASE / XYLUOKINASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U48403; AAC52824.1; -
 DR HSSP; P08859; 1GLJ.
 DR MGD; MGI:106594; GYK.
 DR InterPro; IPR000577; FGGY_kin.
 DR Pfam; PF02370; FGGY; 1.
 DR Pfam; PF02782; FGGY_C; 1.
 DR TIGRfam; TIGR01311; glycerol_kin; 1.
 DR PROSITE; PS00445; FGGY_KINASES_2; 1.
 DR PROSITE; PS00933; FGGY_KINASES_1; 1.
 KM Glycylserol metabolism; Transferase; Kinase; ATP-binding.
 FT NP_BIND 167 179 ATP (PROBABLE).
 SQ SEQUENCE 524 AA; 57458 MW; C3E1C56887683B55 CRC64;

Query Match 11.5%; Score 323.5; DB 1; Length 524;
 Best Local Similarity 25.6%; Pred. No. 4.6e-16;
 Matches 133; Conservative 74; Mismatches 220; Indels 93; Gaps 20;

OY 5 KTVIG-----VDVGGSVRAGIFD-LNGSLSHATEKITTRRSGSVESQEQIMQAVC 58
 DB 6 KAVIGPLVGAVDGTSSTRLFLVNSKTAELSHQVEIKQEPREGWEODPKELDSYV 65
 OY 59 SCIRNA---LTLDVCAQSVAGIGF-DATCSLVVLDK-NGDPLVPSPEGDAKQNIYVMD 113
 DB 66 ECIEKTEKIGQLINDISNIKAIGVSNQRETYVMDKVTGEP-----YAAVWLD 116
 OY 114 HRATEQAE---RINATHHPVLNVGCKISPEMETPKILMKEMPEYERAGO----FF 165
 DB 117 LRQSTVENLSKRIIPGNNNFYKSTGLPLSTYFSAVLRMLDLVKKVOEAVEENRALFG 176
 OY 166 DLADFLWRATGDLARSV-CT--VTCKWTL--AHENRMDDYFRTIGLAELADEDFRI 220
 DB 177 TIDSWLMSLTGIGHGVHCTDYVNASRTMLFNHSLSEMDKELEFEFGIP-----MEI 229
 OY 221 GHHIVSPGTCGNGLTAAQAAEMGLRPTPAVAGLIDAHAGIGTVGEGALNNLAVF 280
 DB 230 LPNVRSSSEIYG-----LMKAGALEGVPLSGCLGDSALVQMGCFODQAKN---TY 279
 OY 281 GTSSCTMAST-----TSPSFVPGVWGYVSAMVPGMLVVGCGSAQAADQL 328
 DB 280 GTGCFELCNHCKVSEHGLLTVAVKLGDKPYVYAL-----EGSVAIAGAVIRWL 332
 OY 329 LDHPFAVEAREMAQRVNOPLPVLADRILEKTAQPSDAVALAKGLHVPEFLGNRAPFA 388
 DB 333 RDNIGITKSEEE-----IEKLAK---EVGTSYGCYFVAFSGLYAPW 372
 OY 389 DPHARAVICGLGMBRDLDNLAYIAGLCIGYGLRQILDQAOTQO-GVNSKNIVISGAG 447
 DB 373 EPSARIGTICGL--TOFTNCHIAIFALEAVCFQFRIELDAMNRDCGIRPLSHLDVGDMT 429
 OY 448 OHPLVROILLADTGIGIPVITTCCEPVLGSAIIGAVAGNI 487
 DB 430 SKKILMQLADLIYIPVKKSPMETTALGAMAAGAEV 469

RESULT 14
 YDEV_ECOLI STANDARD; PRT; 530 AA.
 AC P77432; 099894;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical sugar kinase ydev.
 GN YDEV OR B1511.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubaram S.,
 RA Tagami H., Takeda J., Takemoto K., Takemoto Y., Wada C.,
 RA Yamamoto Y., Horiiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28 0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 182-495 FROM N.A.
 RX MEDLINE-96243037; PubMed-8649811;
 RA Das R., Reddy E.P., Chatterjee D., Andrews D.W.;
 RT "Identification of a novel Bcl-2 related gene, BRAG-1, in human
 RT glioma.";
 RL Oncogene 12:947-951(1996).
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
 CC GLYCOKINASE / XYLUKINASE FAMILY.
 CC CAUTION: WAS THOUGHT BY REF. 3 TO BE A HUMAN SEQUENCE AND WAS
 CC CALLED BY THEM BRAG1 (BRAIN-RELATED APOPTOSIS GENE) (BRAG-1)
 CC WITH A ROLE IN APOPTOSIS. THE DNA SEQUENCE OF THE REGION THEY
 CC SEQUENCED IS MORE THAN 99% IDENTICAL TO THAT OF THIS E. COLI
 CC GENE. FURTHERMORE THEY CLAIM "EXTENSIVE SIMILARITY TO THE
 CC BCL-2 FAMILY OF GENES." SUCH A SIMILARITY IS NOT SIGNIFICANT
 CC AND THIS PROTEIN IS MUCH MORE LIKELY TO BE A SUGAR KINASE.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE000248; AAC74584.1; -
 DR EMBL; D90793; BAA15191.1; -
 DR EMBL; D90794; BAA15198.1; -
 DR EMBL; S82185; AAC17184.1; -
 DR Ecogene; EG31804; ydev.
 DR InterPro; IPR000577; FGGY_kin.
 DR Pfam; PF02370; FGGY; 1.
 DR Pfam; PF02782; FGGY_C; 1.
 DR PROSITE; PS00933; FGGY_KINASES_1; FALSE_NEG.
 DR PROSITE; PS00445; FGGY_KINASES_2; FALSE_NEG.
 KM Hypothetical protein; Transferase; Kinase; Complete proteome.
 FT CONFLICT 490 495 PPBEKH -> TRPGA (IN REF. 2).
 SQ SEQUENCE 530 AA; 57544 MW; C8C3B1E7C8982063 CRC64;

Query Match 11.5%; Score 323; DB 1; Length 530;
 Best Local Similarity 24.4%; Pred. No. 5.1e-16;
 Matches 142; Conservative 74; Mismatches 215; Indels 152; Gaps 26;

OY 7 VIGVDGSGSVRAGIFDNLGSLSHATEK---ITTRRSGSVESQEQIMQAVSCIRN 63
 DB 14 LMALDAGTGSIRAVILEGNCIAVGAEWNRHLAVDPVGS-MEFLNKKMQLACECMRQ 72
 OY 64 ALTLADVCAQSVAGIGFDATCS---LVYLDKNGDPLVPSPEGDAKQNIYV---MDHR 115

Db 73 ALHNAIAGEYIAAV---SACSMREGIYLYNNEGAP-----IMACANVADR 115
 OY 116 ATEOARINATHHPVLN---YVGKISPEMERPKILMLKENPEIERAGOFEDLADFL 171
 Db 116 AAREVELKELNNHTENEYRATGOTLALSAIPRLIMLAHNSDSYIRQASTTMTISDWL 175
 OY 172 TWBRTGLARSVCTVCKMTWLAHENRMPDYEFTIGLAELEADEP-----IRIGHNI 224
 Db 176 AYMLSGELA-----VPSNMGTTGLDLTTTRDMKPALDMMAGLRADI 217
 OY 225 VSGTTCGCGN--GLTAQAAAE--GLPGETPAVAVCLIDAHAG--GIGTV-----GVEGCA-- 272
 Db 218 LSPVKETGTLGLVSSQAELGCKAGTGYVGGVQGLCCGLGVYRPAQTAVLGTFW 277
 OY 273 --LNNLAUYEGTSSCTMASTSPFVGWGPYYSAMVPGTLMVVEGQSAAGAIQGLD 330
 Db 278 QGVVNLA-----APVTDEPMNVN--NPH---VIRGMVAEESISFTTGLTMRFRD 323
 OY 331 FHPAVEE-----AREMAQRVNQPLPVM-----LADRILEKTPQPSDAVALAK 372
 Db 324 AFCAEKEKLAERLIGIDITYLLEEMASRV--PFGSWGVPPIFSDRMRFKT----- 370
 OY 373 GLHVEPELGNRPADPRANAVICGLGMRDLDNLALY-----IAGLCIGYGLR 424
 Db 371 WYHAASFY-----NLSIDPRCKNATLEFALEENAAIVSAC-----NLQ 410
 OY 425 QILDAGTAGVSKNIVISGAGOHPLVROILLADTCGIPVITTTQCCPEVLGSAIIGAVA 484
 Db 411 QILD---FSNIHPSLVFPAFGGSKKLMSSQLADVSGLPNIPVVEKNTALGALAGV 467
 OY 485 GNIAPSVGEAMQOFTVHKYYP-----OERYQSLHHRREYAK 523
 Db 468 AGIFSSMAETGERLVWERTHTPDREKHELVDSDRDKWQAVYQ 510

RESULT 15

XYLB_BACSU
 ID XYLB_BACSU STANDARD: PRT: 499 AA.
 AC P39211:
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Xylulose kinase (EC 2.7.1.17) (Xylulokinase).
 GN XYLB.
 OS Bacillus subtilis.
 OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Borchert S., Klein C., Plksa B., Hammelmann M., Entian K.D.:
 RT "Sequencing of a 26 kb region of the Bacillus subtilis genome
 RT downstream of spoVJ".
 RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertoletti M., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruns C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scottone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weizenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=85297769; PubMed=2994009;
 RA Wilhelm M., Hollenberg C.P.:
 RT "Nucleotide sequence of the Bacillus subtilis xylulose isomerase gene:
 RT extensive homology between the Bacillus and Escherichia coli
 RT enzyme".
 RL Nucleic Acids Res. 13:5717-5722(1985).
 CC -1- CATALYTIC ACTIVITY: ATP + D-xylulose 5-
 CC phosphate.
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
 CC GLYCEKINASE / XYLULOKINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 CC EMBL: 066480; AABA1094.1; -
 CC DR EMBL: 299113; CAB13645.1; -
 CC EMBL: X02795; CA26563.1; -
 CC DR Subtilist; BG10807; xylB.
 CC DR Interpro: IPR000577; FGGT_kin.
 CC DR Pfam: PF02782; FGGY_C; 1.
 CC DR TIGRFAMs: TIGR01312; xylB; 1.
 CC DR PROSITE: PS00445; FGGY_KINASES_2; 1.
 CC DR PROSITE: PS00933; FGGY_KINASES_1; 1.
 CC KW transferase; kinase; xylulose metabolism; Complete proteome.
 CC SEQUENCE 499 AA; 55417 MW; 0772B7E6701E56E CRC64;

Query Match 11.5%; Score 322.5; DB 1; Length 499;
 Best Local Similarity 23.6%; Pred. No. 5,1e-16;
 Matches 130; Conservative 93; Mismatches 241; Indels 87; Gaps 20;
 OY 5 KTVIGDVSGSVRAGIFDNGSLSHATEKITTTTRSGSRVBOSSQELIQAASCIRNA 64
 Db 2 KYVIGDITGSAVKTLIVNONGKVCATSKRYPLDIOKACYSQNPEDWQITETALAE 61
 OY 65 LTLADYCAQVAIGDPAIC-SLVVLDKNGDPLPYSPBEGAKONITVMMHRTBOAERI 123
 Db 62 VSIISNQAOKIDIDISYSGNHGLVLLDDOQVLA-----RNAILMNTTRTPPCIRN 112
 OY 124 N---ATHNPLVNVGKISPEMERPKILMLKENPEIERAGOFEDLADFLVTRATGDLA 180
 Db 113 TEKFGDH--LDDITKRNVLGFLPKMLWKHNEPELFKTAFLPKVDYRRRMGVIN 170
 OY 181 RSVCTVCKMTWLAHENR--WDDPYFTTGLAELEADEFTIRGHNIYSP--GTPCGNGL 235
 Db 171 TEVSDA--GTLHLHTRKEMSNIDICQIGIS--AD-----ICPLVESHCVGSL 217
 OY 236 TAAAAEMGLLPTPAVAVGLIDAHAGIGTVGEGCALNNLAVFGTSSCTMASTSPS- 294
 Db 218 LPHVAAKTGLETKYVAGADNACAGTAGILSSG--KTLCSI-GTSGVILSYEEKER 274

```

0Y 295 -----FVGVGVGVGYSAWVPGMLVVGQSGAAGAIQOLDDPHRPAVEAREMQRVN 346
Db 275 DEKGVHVFHNFGKKDSFTM-----GYTLAAGISLMF-----KRTF 311
0Y 347 OPLPWLADRLLEKTPAOPSDAVAL-AKGLHAVPEELGNRAPFADPHARAVIGSIGMERDI 405
Db 312 AP-----NESFEOLQGVENIPIGANGLLYTPYLVGEPRPHADSIRGSLIGMGHNH 365
0Y 406 DNLALYTAGCGIGYGRQLDQAOTQGVYVSKNIVISGGAGOHPLVROLAOTCGAPY 465
Db 366 KHFLLR--ALMEGTTFSLHSEIELFRAGKSVHTIVSISGGAKNDYTLQAOADIFNTRYI 422
0Y 466 TTQCCPEVLLGSAILGAVAGNIADSVGEAMQOETHVDKYYYPO---ERYQSLHRRREYA 521
Db 423 KLENQGGPAMGAAMLAAFGSGMFESLEBCAEQFIREAAAEVAPKAQNVQYKTKTL---FDL 478
0Y 522 YKOL-QHTAKL 531
Db 479 YKNITYHTKDL 489

```

Search completed: March 13, 2003, 16:53:44
Job time : 12.6093 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:43:37 ; Search time 34.0381 Seconds

(without alignments)
3232.531 Million cell updates/sec

Title: US-09-802-208B-4

Perfect score: 2811
Sequence: 1 MTITKTVIGDVSGSVAG.....HHRRYEAYKOLQHTAKLRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp_bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2811	100.0	534	2	09F4L6	09F4L6 escherichia
2	2031	72.3	535	2	052716	052716 klebsiella
3	1666	59.3	536	16	08U7X5	08U7X5 agrobacteri
4	1560.5	55.5	538	16	08YBC1	08YBC1 bruceella me
5	1244.5	44.3	545	16	08ZAZ2	08ZAZ2 yersinia pe
6	1167.5	41.5	569	10	09MOC9	09MOC9 arabidopsis
7	1061	37.7	528	16	098D08	098D08 rhizobium l
8	1050	37.4	527	16	092N09	092N09 rhizobium m
9	1005.5	35.8	548	5	09VZJ8	09VZJ8 drosophila
10	971	34.5	525	16	08U7J6	08U7J6 agrobacteri
11	940	33.4	439	4	096C11	096C11 homo sapien
12	887	31.6	598	3	096UC2	096UC2 neurospora
13	864.5	30.8	715	3	004585	004585 saccharomyc
14	769.5	27.4	439	5	095T84	095T84 drosophila
15	672	23.9	246	4	09HA63	09HA63 homo sapien
16	465	16.5	509	16	092U15	092U15 rhizobium m

17	447	15.9	252	11	09D7H0	09D7H0 mus musculus
18	434	15.4	252	4	09NV20	09NV20 homo sapien
19	407	14.5	500	2	030364	030364 thermomanaer
20	381.5	13.6	500	16	097FW4	097FW4 clostridium
21	368.5	13.1	530	16	08Z1S4	08Z1S4 yersinia pe
22	362	12.9	492	16	09WXX1	09WXX1 thermotoga
23	353	12.6	530	16	08Z2X3	08Z2X3 salmonella
24	352	12.5	530	16	08ZK06	08ZK06 salmonella
25	339.5	12.1	522	16	09CLG3	09CLG3 pasteurjella
26	336	12.0	493	17	028522	028522 archaeoglob
27	334.5	11.9	502	16	0911D5	0911D5 pseudomonas
28	328.5	11.7	504	16	0927C9	0927C9 listeria in
29	326	11.6	504	16	08XAY5	08XAY5 escherichia
30	324	11.5	501	16	09RT38	09RT38 deinooccus
31	322.5	11.5	509	16	092M66	092M66 rhizobium m
32	320	11.4	504	16	08Y3W7	08Y3W7 listeria mo
33	319.5	11.4	511	16	08U9Y5	08U9Y5 agrobacteri
34	319	11.3	494	16	09HY41	09HY41 pseudomonas
35	318	11.3	497	16	08R8J4	08R8J4 thermomanaer
36	316	11.2	506	16	09WYC0	09WYC0 thermotoga
37	316	11.2	513	17	09YCG4	09YCG4 aeropyrum p
38	314.5	11.2	519	16	09H2T2	09H2T2 pseudomonas
39	312.5	11.1	501	17	097XW1	097XW1 sulfolobus
40	312.5	11.1	515	16	092M02	092M02 rhizobium m
41	312	11.1	494	16	08R8Q3	08R8Q3 thermomanaer
42	310	11.0	477	16	092MP4	092MP4 rhizobium m
43	310	11.0	493	2	030495	030495 pseudomonas
44	309.5	11.0	484	16	08ZL91	08ZL91 salmonella
45	309.5	11.0	484	16	08Z2B7	08Z2B7 salmonella

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	534 AA.
ID 09F4L6			
AC 09F4L6			
DT 01-MAR-2001 (TREMBLrel. 16, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Ribitol kinase.			
GN RTLK.			
OS Escherichia coli.			
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC Escherichia			
OX NCBI_TaxId=562;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=C;			
RA Lafayette P.R., Parrott W.A.			
RT "A non-antibiotic marker for amplification of plant transformation			
RT vectors in E. coli";			
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR EMBL: AY005817; AAC01884.1; -			
DR InterPro: IPR000577; FGGY_kin.			
DR Pfam: PF00370; FGGY_1.			
DR Pfam: PF02782; FGGY_C_1.			
DR TIGRPFAMs: TIGR01315; 5C_CHO_kinase; 1.			
KW kinase.			
SQ SEQUENCE 534 AA; 57343 MW; F62DD4B966556875 CRC64;			
Query Match	100.0%;	Score 2811;	DB 2; Length 534;
Best Local Similarity	100.0%;	Pred. No. 2.7e-187;	
Matches 534;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
OY 1 MTITKTVIGDVSGSVAGIFDLNGLSLSHATEKITTTTTRSSGRVYQSSQETMQAVCSC 60			
DB 1 MTITKTVIGDVSGSVAGIFDLNGLSLSHATEKITTTTTRSSGRVYQSSQETMQAVCSC 60			
OY 61 IRNALTLADYCAOSVAGIGDATCSLVLDKNGDPLVPSPEGAKONIIYMDHRTAEQA 120			
DB 61 IRNALTLADYCAOSVAGIGDATCSLVLDKNGDPLVPSPEGAKONIIYMDHRTAEQA 120			

QY 121 ERINATHHNRYLVNKGKISPEMERPKILMLKENNPEIYERAGOFFDLADFLTWRTATGDLA 180
 DB 121 ERINATHHNRYLVNKGKISPEMERPKILMLKENNPEIYERAGOFFDLADFLTWRTATGDLA 180
 QY 181 RSVCTVCKMTWLAHNRMPDVFRTGTLAEADDEFIRIGHIYVSGTFCGNGLTQAQA 240
 DB 181 RSVCTVCKMTWLAHNRMPDVFRTGTLAEADDEFIRIGHIYVSGTFCGNGLTQAQA 240
 QY 241 AEMGLPCTPVAVGLIDAHAGIGTGVGEGALNNLAVFEGTSSCTMASTSPSPFVPGVW 300
 DB 241 AEMGLPCTPVAVGLIDAHAGIGTGVGEGALNNLAVFEGTSSCTMASTSPSPFVPGVW 300
 QY 301 GPYISANVPGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 360
 DB 301 GPYISANVPGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 360
 QY 361 TAQPSDVAVALAKGLHVPPEFLGNRAPFADPHARAVICGLGERLDNLALYTAGLGIGT 420
 DB 361 TAQPSDVAVALAKGLHVPPEFLGNRAPFADPHARAVICGLGERLDNLALYTAGLGIGT 420
 QY 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLLGSAIL 480
 DB 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLLGSAIL 480
 QY 481 GAVAGNTAPSVGEAMQOFTHVDKYYPQERYOSLHRRREAYKOLQHTAKLLRD 534
 DB 481 GAVAGNTAPSVGEAMQOFTHVDKYYPQERYOSLHRRREAYKOLQHTAKLLRD 534

RESULT 2

ID 052716 PRELIMINARY; PRT; 535 AA.
 AC 052716;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Ribitol kinase.
 GN RBRK.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-1033-5P14;
 RC MEDLINE=97464425; PubMed=9324246;
 RA Heuel H., Turgut S., Schmid K., Lengeler J.W.:
 RA "Substrate recognition domains as revealed by active hybrids between
 RT the D-arabinitol and ribitol transporters from Klebsiella
 RT pneumoniae";
 RT J. Bacteriol. 179:6014-6019(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-1033-5P14;
 RP MEDLINE=98304087; PubMed=9639934;
 RA Heuel H., Shakeri-Garakani A., Turgut S., Lengeler J.W.:
 RA "Genes for D-arabinitol and ribitol catabolism from Klebsiella
 RT pneumoniae";
 RT Microbiology 144:1631-1639(1998).
 DR EMBL: AF043244; AAC26495.1;
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR Pfam: PF02782; FGGY_C1.
 DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
 KW kinase.
 SQ SEQUENCE 535 AA; 56795 MW; A5C81E4CC2F2B7F9 CRC64;

Query Match 72.3%; Score 2031; DB 2; Length 535;
 Best Local Similarity 72.3%; Pred. No. 4,7e-133;
 Matches 384; Conservative 49; Mismatches 98; Indels 0; Gaps 0;

QY 4 TKYIVGVDSGSVRACIFDLNGLSLSHATEKITTTTRRSGRSVEQSSQEIWQAVCSICRN 63

DB 5 TONTIGVDVSGSVRAGFNLRGELAHAFREITLFRAGAKRVQSSREIWOAVCYCIKT 64
 QY 64 ALTLADVCAQSVAGIGFADATSLVYLDKNGDPLVPSPEGDAKONIYIMDHRAEQAEI 123
 DB 65 AVANAQVSPSSIAIGIDATCSLVYIGDNPAPLAVGSDDDAKRIIYIMDHRAEQAEI 124
 QY 124 NATHHPVNVNKGKISPEMERPKILMLKENNPEIYERAGOFFDLADFLTWRTATGDARSV 183
 DB 124 NATHHPVNVNKGKISPEMERPKILMLKENNPEIYERAGOFFDLADFLTWRTATGDARSV 183
 QY 184 CTYVCKMTWLAHNRMPDVFRTGTLAEADDEFIRIGHIYVSGTFCGNGLTQAQAEM 243
 DB 184 CTYVCKMTWLAHNRMPDVFRTGTLAEADDEFIRIGHIYVSGTFCGNGLTQAQAEM 243
 QY 244 GLLTGPVAVGLIDAHAGIGTGVGEGALNNLAVFEGTSSCTMASTSPSPFVPGVW 303
 DB 245 GLLTGPVAVGLIDAHAGIGTGVGEGALNNLAVFEGTSSCTMASTSPSPFVPGVW 304
 QY 304 YSAMPVGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEKTAQ 363
 DB 305 YSAMPVGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEKTAQ 364
 QY 364 PSDAVVALAKGLHVPPEFLGNRAPFADPHARAVICGLGERLDNLALYTAGLGIGT 423
 DB 365 PSEAVTLAAGLHVPPEFLGNRAPFADPHARAVICGLGERLDNLALYTAGLGIGT 424
 QY 424 RQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLLGSAILGAV 483
 DB 425 RQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLLGSAILGAV 484
 QY 484 AGNAPSVGEAMQOFTHVDKYYPQERYOSLHRRREAYKOLQHTAKLLRD 534
 DB 485 AGNAPSVGEAMQOFTHVDKYYPQERYOSLHRRREAYKOLQHTAKLLRD 535

RESULT 3

ID 0807X5 PRELIMINARY; PRT; 536 AA.
 AC 0807X5;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Ribitol kinase.
 GN AYU4324 OR AGR_L1075.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-G., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.:
 RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58";
 RT Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S.,
 RA Guorillo B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wolian C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gutson J., Lomo C., Seac C., Strub G.,

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gotzelli J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Splet E., Spadling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003480; AAF47823.1;
 DR FlyBase: FBgn003484; CG11594.
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
 DR SEQUENCE 548 AA; 59641 MW; D934CA8D72A0E6B CRC64;
 SQ
 Query Match 35.8%; Score 1005.5; DB 5; Length 548;
 Best Local Similarity 39.4%; Pred. No. 9.6e-62;
 Matches 215; Conservative 83; Mismatches 216; Indels 31; Gaps 6;

DB 186 CKMNYDAANGSMNKEFLKQADLEELTONNFEKLSGVOPGRTVYKGTATKAAAGELISA 245
 QY 248 GTPAVAGLIDAHAGGIGTVG-----VEGALNNLAVYVGTSCSTMASTSPSEFV 296
 DB 246 GTVAVSTSLIDAHAGALMGCGRSKESKSGADVDG-----KMLLIAGTSCCHMSTRKACFA 301
 QY 297 PGWMPYYSAMVPGILMVEGSGAAGAAIDQLDFHPAVEARE-----MAQRVNOPL 349
 DB 302 QGVWGPYODALIPGVFLNEGSGIAGHLDHVLSHESYAEALQSOLGDEKFIYOHNLNLL 361
 QY 350 PWNLAADRLTEATAPSDVAVLAKGLHVPPELGNRAPADPHARAVICGEMEDDNL 409
 DB 362 PELAAARL-----SOVGLTDVHWVPDHNKNSPIADPTLKGVTGLDMTGTESLA 415
 QY 410 ALVYAGLIGIGYGRLOIDAQTAOG-VVSKNIVISGAGHPVROLIADTCGIPVITQQ 468
 DB 416 IKYLAFLVAGLAVGRHIENLYOGRAPFOFLLECGGLAKNPVYGCADICNLPALIPD 475
 QY 469 CCEPVLLGSAILGAVAGNIAPSVGBAMQOFTYVDKYYRPOERYSLHRRYEAUKQLOHT 528
 DB 476 EOEWVLGAAALGAASGHPDLSASKSMGQTOLVKNPNETLEFHNRYKVEFLOLLEN 535
 QY 529 AKLR 533
 DB 536 QROTR 540
 RESULT 10
 ID 080736 PRELIMINARY; PRT; 525 AA.
 AC 080736;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DE Ribitol kinase.
 GN AT04453 OR AGR_L_826.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
 RA Chen Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Oken V., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kuyavian T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
 RA Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009374; AAL45247.1;
 DR EMBL: AE008240; AAK88987.1;
 KM Kinase: Complete proteome.
 SQ SEQUENCE 525 AA; 56072 MW; 0C96C97306BED7C0 CRC64;


```

QY 176 TGDARSVCITCKWTL-----AHENKMDPDYFRTIGLAELADEDFIRIG-----HHI 224
D 186 TGKREIFSCSTVCKQKQGVGVDSVKGWQDEYFETIGLGLTIDNFIKMGVGVNGSW 245
QY 225 VSPETPCNGITLQAAMEMGLLPCTPAVAGLIDAHAGIGTVGE-----269
D 246 MSAGELVGG-GLSEAGGQGLIPAGIAGISGVIDYAGWIGVAKVGLSPDRDSDVAPN 304
QY 270 --GGLANLAVVFGTSSCTMASTSPSFVPGVWGPYSAVPGMLVGEQSGAAGAIIDQ 327
D 305 DVSGAFRLAAVAGTSTCHLAMSNGPVFPGWGPYRNVLLIPGYMAEGGSAIGELMKH 364
QY 328 LLDHRAVEEREMAQRYNOLPWMLADRIE---KTAQPSDAVALAKGLHVPFLGNR 384
D 365 MLEHHVAYDETVKAEAGAKNIYDYLNEHLEKMAKKTAPSIY-LVHFFFGYGLDMGNR 423
QY 385 APFADPARAVYICGLGEMRDNLALYIAGLIGYGLRGLDQAOTAGVSKNIVISG 444
D 424 SPIDPMKRGATIGMSNDKSGOMALLYSTMERIALOTROIIVEMNTAGHTISFMSG 483
QY 445 GAGHPPLVROLADTCGIPVITTOCE--PVLLGSAIIGAVGNIAPS-----VGEAMQF 498
D 484 SQCNELIMDLIATACDMPVLIPIRYVNAAYVGAAMLGAKAASAKPOSTTEPLMDINDRM 543
QY 499 THVDKYYT---PQERTQSLHRRYEAYKOLQHTAKLLR 533
D 544 SKPGKTVMSRGDPAEK--KLDAKYEIIFLDARTQOEYR 580

```

RESULT 13

004585 PRELIMINARY; PRT; 715 AA.

```

ID 004585
AC 004585;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 79.2 kDa protein.
GN YDR109C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Murphy L., Shore L., Harris D.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrett B., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48758; CA88663.1; -.
DR SGD; S0002516; YDR109C.
DR InterPro; IPR000577; FCGY_kin.
DR Pfam; PF00370; FCGY_1.
DR TIGRFAMs; TIGR01315; 5C_CHO_kinase; 1.
KW Hypothetical protein.
SQ SEQUENCE 715 AA; 79160 MW; A52B2D9B4B4C9C7E CRC64;

```

Query Match 30.8%; Score 864.5; DB 3; Length 715;
 Best Local Similarity 36.1%; Pred. No. 9.1e-52;
 Matches 201; Conservative 100; Mismatches 211; Indels 45; Gaps 10;

```

QY 5 KTVGVGVGSSVAGIFDNGSLSHATEKITTRSGSVHEQSSQEIQAVGSCIRNA 64
D 38 KFYGVGVVGGISARACYIDSGNMLSLAEKTKRQILSNFTOSSREIMAAVCYQRTV 97
QY 65 LTLADVCAQSVAGIGFATSLVLD-KNGDPLVSPF-GDAKONITVMDHRAETQDAER 122
D 98 VEESGVDPERVRIGIGFATSLVYVSATNFEIAGVDFTNNDONITIMDHRAKETEER 157

```

```

QY 123 INATHHPVLNVYGGKISPEMETPKIILKENMPEIYERAGOFEDLADFLTRATGLARS 182
D 158 INSSGDKCLKLVGGOMSVEMEIPKIKMLKNLEAGIFQDCKFPLDPLYLFKAKGKENS 217
QY 183 VCTVCKWTL-----AHENKMDPDYFRTIGLAELADEDFIRIG-----HHIYSPGPC 231
D 218 FCSAVCKQGLFVGVGSDIGMSKEFLNSIGLSELTKNDFERLGLSLREKKNFILA-EC 276
QY 232 GNGLTQAAMEMGLLPCTPAVAGLIDAHAGIGTVGE-----GGL 273
D 277 ISPLDKKAQGLGLEHCVSSGIIDYAGWGVAAKPEAVAGLAETENYKKDFNGAI 336
QY 274 NMLAVFGTSSCTMASTSPSFVPGVWGPYSAVPGMLVGEQSGAAGAIIDQLDHP 333
D 337 GRLLAVAGTSTCHILSKNPJFVHGVGPPYDVLARGLMAEGQSGCVGLDLITTHP 396
QY 334 AVEEREMAQRYNOLPWMLADRIEKTAP---SDAVALAKGLHVPFLGNRAPFADP 390
D 397 AFTELSHMANLAGSKFEYL-NKILETLVERKRYRVSLSLKLHFFGYDIGNSPINDP 455
QY 391 HARAIVICGLGEMRDNLALYIAGLIGYGLRGLDQAOTAGVSKNIVISGAGQHP 450
D 456 NMRACIIQSMDNSIEDLAVYLSACEFISQOTROIIEVMLKSGHEINAIEMSGQCRNS 515
QY 451 LVROLADTCGIPVITTOCE--PVLLGSAIIGAVGNIAPSVGEAMQFTHVDKYYTPOE 509
D 516 LMRLLADCTGLPIVIRPYDAAVFGSALGAAASDE-----DYTRKRTLKGOK 567
QY 510 RYQSLHRRYEAYKOLQ 526
D 568 SSQTKTERFNDYSISQ 584

```

RESULT 14

095T84 PRELIMINARY; PRT; 439 AA.

```

ID 095T84
AC 095T84;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE G412991P.
GN CG11594.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Abdayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060286; ALU25325.1; -.
DR FLYBase; FBgn0035484; CG11594.
DR InterPro; IPR000577; FCGY_kin.
DR Pfam; PF02782; FCGY_C; 1.
DR TIGRFAMs; TIGR01315; 5C_CHO_kinase; 1.
SQ SEQUENCE 439 AA; 48172 MW; F26903DBB321A40E CRC64;

```

Query Match 27.4%; Score 769.5; DB 5; Length 439;
 Best Local Similarity 37.9%; Pred. No. 1.8e-45;
 Matches 167; Conservative 69; Mismatches 176; Indels 29; Gaps 5;

```

QY 112 MDRATQDAERINATHHPVLNVYGGKISPEMETPKIILKENMPEIYERAGOFEDLADFL 171
D 1 MDHRAEQETQIRINAFKSLKLVYGGQVSLMEVYFKLMLKNLSQTFGINIRVDFLDL 60
QY 172 TWRTATGLARSVCITCKWTLAHENKMDPDYFRTIGLAELADEDFIRIGHIYSPGPC 231

```

```

Db 61 TWRATGVDRSLCSVYCKWNYDANGSMWNEFLKQADLEELTONNEKLGSDVQPPGRIV 120
OY 232 GNGLTQAQAAEMGLPPTPAVGLIDAHAGIGTVG-----VEGALNNLAYVE 280
Db 121 GKCLTKAKAGELGSLGTVYSTSLIDHAGALMFGRSKESGADVDG---KKAFLA 176
OY 281 GTSSCTMASTTSPSPVPGWGPYYSAMVPGMLVEGQSAAGAIDOLLFHPHAEARE 340
Db 177 GTSTCHMSITRKACFGAGWGPYODALIRPGYFNEGQSTAGLLDHVLSKSHSYAEIKS 236
OY 341 -----MAQRVQPLPWLADRILEKTAQPSDAVALAKGLHVPPEFLGNRAPDPHAR 393
Db 237 QLGEDKFYOHNLKLLPELAAARGL-----SQVGCLTDVHWVPLDGNRSPIDPTLR 290
OY 394 AVICGIGMERDLDNLALYTAGLCIGYGLRQLDPAOTAG--VYSKNIVISGAGOHPLV 452
Db 291 GVTITGLDMTRGTESLAIKTYLAFVQALAYGTRHLEINLYOYGRAPFQTLFCGGLAKNPLY 350
OY 453 RQIADTCGIPYITTOCCEPVLGSAILGAVAGNIAPSVGEAMQOFTHVDRYXYRQERYQ 512
Db 351 VQCHADICNLPALIPDEQENVLVGAALGAAGHFDLSLEASKSMGTQLVKPNAETL 410
OY 513 SLHHRREYAVKOLQHTAKLR 533
Db 411 EFHNRKRYKVFQLLENQRYR 431

```

RESULT 15

```

O9HA63 PRELIMINARY; PRT: 246 AA.
ID 09HA63
AC 09HA63;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ12175 f1s, Clone MAMMA1000713, weakly similar to
DE L-ribulokinase (EC 2.7.1.16).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 11)
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022237; BAB13993.1; -
SQ SEQUENCE 246 AA; 26576 MW; 04D5374635B4DB01 CRC64;

```

Query Match 23.9%; Score 672; DB 4; Length 246;

Best local Similarity 52.0%; Pred. No. 4, 8e-39;

Matches 127; Conservative 38; Mismatches 67; Indels 12; Gaps 4;

```

OY 110 VMMDHRAETGAERINATNHPVLNVYGGKISPEMETPKIIMLKEMPEI-YERAGQFFDLA 168
Db 1 MMIDHRAVSQVNRINETHKHSVLQYGVGMSVEQAPKLLMKELREICWDKAGHFDLP 60
OY 169 DELTPRATGDLARSVCYVCKKWTMLAHENRWDYFRTIGLAELEDFIRIGHIIVSPG 228
Db 61 DFLSMKATGYTARSCLSVCKWITYSA-EKGMWDSFWMKMIGLEDFVADNYSKIGNOVLPG 119
OY 229 TPCGNGLTQAQAAEMGLPPTPAVGLIDAHAGIGTVG--VEGAL-----NNLAY 278
Db 120 ASIGNGLTPEARADLGLIPGIAVAASLIDHAGLGIVGADVGRHGLICGQPTSKLAY 179
OY 279 VFGTSSCTMASTTSPSPVPGWGPYYSAMVPGMLVEGQSAAGAIDOLLDFHPHAEAE 338

```

```

Db 180 ICGTSSCHMGISKDPIFVPGWGPYFSAMVPGFWLNEGOSVYGLIDHNVQGHAAFPPEL 239
OY 339 REMA 342
Db 240 QVKA 243

```

Search completed: March 13, 2003, 16:55:12
Job time : 37.0381 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 ; Search time 34.7483 Seconds
(without alignments)
1625.927 Million cell updates/sec

Title: US-09-802-208B-5
Perfect score: 2211
Sequence: 1 MSRRNKQWIGLPLHLWGVI.....RVQPGFSSAPVTEKALNIS 424

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries:

Database : A.Geneseq.101002.*
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2211	100.0	424	22	AAE09781
2	1040.5	47.1	467	23	ABP38175
3	580.5	26.3	445	22	AA689873
4	174.5	7.9	414	20	AAV37671
5	168.5	7.6	457	21	AAAB18235
6	164	7.4	394	22	ABBA47287
7	158.5	7.2	387	23	ABBS3425
8	158.5	7.2	416	20	AAV02610
9	158.5	7.2	416	22	AA898023
10	158.5	7.2	416	22	AA84207

11	155	7.0	397	23	ABP28804
12	152.5	6.9	392	22	AAE05852
13	147	6.6	401	23	ABBA4816
14	145	6.6	393	23	ABP40749
15	143.5	6.5	484	22	AA876735
16	143.5	6.5	1595	22	AA693189
17	142	6.4	497	22	ABBB60537
18	141	6.4	352	20	AAV35284
19	141	6.4	496	23	ABBS4015
20	140	6.3	447	22	AA84873
21	139	6.3	417	22	AAU34999
22	138.5	6.3	449	22	AA692801
23	138	6.2	388	18	AAW26740
24	138	6.2	454	22	ABBS2906
25	137	6.2	382	23	ABBA7392
26	137	6.2	405	23	ABBS3977
27	136.5	6.2	425	22	AAU34874
28	136.5	6.2	425	22	AA698958
29	136	6.2	492	23	ABP40002
30	133.5	6.0	490	23	ABP30110
31	132.5	6.0	378	22	AA692975
32	132.5	6.0	402	22	AA876763
33	132.5	6.0	412	22	AA878973
34	132.5	6.0	423	23	ABP40365
35	132	6.0	430	20	AAW88344
36	131	5.9	433	23	ABP27912
37	131	5.9	466	22	ABBS2555
38	130.5	5.9	467	22	ABBS6687
39	130.5	5.9	542	22	ABBS6844
40	130	5.9	518	22	AA835474
41	130	5.9	701	22	AA692442
42	129.5	5.9	453	22	AA693123
43	129	5.8	400	23	ABP27911
44	129	5.8	400	23	ABP29844
45	129	5.8	401	21	AAV91291

ALIGNMENTS

RESULT 1
AAE09781
ID AAE09781 standard; Protein; 424 AA.
XX
AC AAE09781;
XX
DT 29-NOV-2001 (first entry)
XX
DE Escherichia coli strain C rfl operon encoding ribitol transporter.
XX
KW Positive selection system: metabolic; arabinol; ribitol; mannitol;
KW transgenic cell; marker gene; ribitol dehydrogenase; ribitol kinase;
KW ribitol transporter; rfl operon.
XX
OS Escherichia coli C.
XX
PN WO200166779-A2.
XX
PD 13-SEP-2001.
XX
PE 08-MAR-2001; 2001WO-US07474.
XX
PR 08-MAR-2000; 2000US-0188291.
PR 15-AUG-2000; 2000US-0255595.
XX
(UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Parrott W, Lafayette P, Kane P;
XX
DR WPI: 2001-565596/63.
DR N-PSDB: AADI6811.
XX
PT Positively selecting transformed cells comprising selectable marker

Streptococcus poly
Pseudomonas stutzeri
Listeria monocytogenes
Staphylococcus epi
Corynebacterium g1
C glutamicum prote
Drosophila melanog
Chlamydia pneumonia
Lactococcus lactis
Lactococcus lactis
Enterococcus faecis
C glutamicum prote
Staphylococcus car
Escherichia coli p
Listeria monocytogenes
Lactococcus lactis
E. coli cellular p
E. coli growth and
Staphylococcus epi
Streptococcus poly
C glutamicum prote
Corynebacterium g1
C. glutamicum SRT
Staphylococcus epi
Salmonella enteric
Streptococcus poly
Escherichia coli p
Lactococcus lactis
Drosophila melanog
B cepacia multidru
C glutamicum prote
C glutamicum prote
Streptococcus poly
Streptococcus poly
Group B Streptococcus

PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinol, ribitol which confer selective advantage on transformed
PT cells
XX
XX Claim 18; Page 37; 37pp; English.
CC The present invention relates to a positive selection system that
CC involves conferring to transferred cells the ability to metabolise
CC arabinol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C ribitol transporter encoded by ribitol operon (rtl operon). The
CC operon also encodes ribitol dehydrogenase and ribitol kinase.
CC
XX
SQ Sequence 424 AA:
Query Match 100.0%; Score 2211; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 3.3e-227;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRRNKMWLPRLHNGYIAIAVFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLA 60
DB 1 MSRRNKMWLPRLHNGYIAIAVFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLA 60
QY 61 AALSAMWSGVVAEITTPQKTMKLGFLVLCVFNHFLVFLVGLQANYGILLFYGIRGLAYP 120
DB 61 AALSAMWSGVVAEITTPQKTMKLGFLVLCVFNHFLVFLVGLQANYGILLFYGIRGLAYP 120
QY 121 LFLYSFIVYIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
DB 121 LFLYSFIVYIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
QY 121 LFLYSFIVYIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
DB 121 LFLYSFIVYIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
QY 181 FCFPGGYIAIISLRHVTPGHHNLTPREKPAELSRVATLLYTRNFLLSIVITNTLS 240
DB 181 FCFPGGYIAIISLRHVTPGHHNLTPREKPAELSRVATLLYTRNFLLSIVITNTLS 240
QY 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTTIFSNIFWGIVAEKMGMRVIRMGCL 300
DB 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTTIFSNIFWGIVAEKMGMRVIRMGCL 300
QY 301 GMAASSLAFFYIMPOYFGHNWMMAMIPAIAGTFVAAEFVPMMAVFPALPEPKHGAISYVN 360
DB 301 GMAASSLAFFYIMPOYFGHNWMMAMIPAIAGTFVAAEFVPMMAVFPALPEPKHGAISYVN 360
QY 361 LSAGMSNFLPAIVVLLPWEITIGVIAVATLLAFVLCARIVPOPGSSAPVTEKA 420
DB 361 LSAGMSNFLPAIVVLLPWEITIGVIAVATLLAFVLCARIVPOPGSSAPVTEKA 420
QY 421 LNTIS 424
DB 421 LNTIS 424
RESULT 2
ABP38175
ID ABP38175 standard; Protein; 467 AA.
XX
XX ABP38175;
XX
XX 24-JUL-2002 (first entry)
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3020.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KM antibacterial; gene therapy.

XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 9805-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI, 2002-381255/41.
XX
XX N-PSDB; ABN90720.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
XX Disclosure; SEQ ID 3020; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
SQ Sequence 467 AA:
Query Match 47.1%; Score 1040.5; DB 23; Length 467;
Best Local Similarity 48.6%; Pred. No. 3.3e-102;
Matches 195; Conservative 76; Mismatches 129; Indels 1; Gaps 1;
QY 10 GLPLHLIMGYIAIAVFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLAALISAVSG 69
DB 55 GLPLSLIMGFVAIAVFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLAALISAVSG 114
QY 70 VVAETTPQKTMKLGFLVLCVFNHFLVFLVGLQANYGILLFYGIRGLAYPLFYSTIY 129
DB 115 VLAETFGRRIMLIGVAAIVFHVFLFLGLEQOKYTLMAVFGIRGLAYPLFYATMWW 174
QY 130 IHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALAFCPAGGVIA 189
DB 175 IAGVTPPARMAASANGMWSMTCIGLIGLKNMIPSLISRGIFINTLWGVFWVAALMI 234
QY 190 MISLRHVTPGHHNLTPREKPAELSRVATLLYTRNFLLSIVITNTLSLGFAYIMP 249
DB 235 MYLWKEGAGKPPDAVTLVERLKLTLSSGVITIIARRRGMPFLIVIRICNLSLFGPLPILP 294
QY 250 MMFVD-ELGFTTSEMLQVMAAFETTTIFSNIFWGIVAEKMGMRVIRMGCLGMAASSLA 308
DB 295 LYLTSVEVGFSSMEQMLHMGVMEFVVSLETVNIWQJSDRLGMLQMWKPGICGALSLSLA 354
QY 309 FYVMPQYFGHNWMMAMIPAIAGTFVAAEFVPMMAVFPALPEPKHGAISYVNLSAGMSNF 366
DB 355 FYVLPQYGAHFELALIAIGFGISVAFPMGAVFLALAPDEKGAISAHNLAAAGLSNF 414
QY 369 LAPAIIVVLLPWEITIGVIAVATLLAFVLCARIVNEOP 409
DB 415 MGPGIATLTATLTGIGKGVWVIYAGLVAVAGAVLFFIIVPOP 455
RESULT 3
AAG89873

ID	AA089873	standard; Protein; 445 AA.
XX	AA089873	
AC	AA089873	
DT	26-SEP-2001	(first entry)
DE	C glutamicum protein fragment SEQ ID NO: 3627.	
XX		
KW	Corynebacterium; amino acid synthesis; vitamin; saccharide; organic acid synthesis.	
XX		
OS	Corynebacterium glutamicum.	
PN	EP1108790-A2.	
PD	20-JUN-2001.	
PF	18-DEC-2000; 2000EP-0127688.	
XX		
PR	16-DEC-1999; 99JP-0377484.	
PR	07-APR-2000; 2000JP-0159162.	
PR	03-AUG-2000; 2000JP-0280988.	
XX		
PA	(KYOW) KYOWA HAKKO KOGYO KK.	
PI	Nakagawa S, Mizoguchi H, Ando S, Haysashi M, Ochiai K, Yokoi H;	
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;	
XX		
DR	WPI; 2001-376931/40.	
XX		
DR	N-PSDB; AAH65092.	
XX		
PT	Novel polynucleotides derived from coryneform bacteria, for identifying	
PT	mutation point of a gene, measuring expression of a gene, analysing	
PT	expression profile or pattern of a gene and identifying homologous gene	
XX		
PS	Claim 17; SEQ ID NO: 3627; 246pp + Sequence Listing; English.	
XX		
CC	The present invention provides a number of nucleotide and protein	
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These	
CC	are useful for identifying the mutation point of a gene derived from a	
CC	mutant of coryneform bacterium, measuring expression amount and	
CC	analysing the expression profile or expression pattern of a gene derived	
CC	from Coryneform bacterium, and identifying a homologue of a gene derived	
CC	from coryneform bacterium. Coryneform bacteria are useful for producing	
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,	
CC	particularly L-lysine. The present sequence is a protein described	
CC	in the exemplification of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from the	
CC	European Patent Office.	
XX		
XX		
SO	Sequence 445 AA;	
Query Match	26.3%; Score 580.5; DB 22; Length 445;	
Best Local Similarity	30.4%; Pred. No. 3.8e-53;	
Matches 131; Conservative	89; Mismatches 188; Indels 23; Gaps 8;	
9	IGLPHLIMGYAIAVPMFGDGEFLAFLSHYKRS-LGFTPAASFAFTLYGLAALAMV 67	
12	LGIPRLPIGFGLTFTMGDVENNILEPFLSSHGFSVSLAGTLVTVYGAVALAAAF 71	
68	SCVVAEITTPOKTMLIGFLMCVFHVLFLVFCAGANYGLLLEFGIRGLAPLELFSFI 127	
72	AAALSDIMGPKRVMILGASIWIVFELIFLTVALTTDHTPLIFLANGLRGFGFPFAYGFL 131	
128	VIIHNVSSENSSALGWYAVVYSGVIGVAGSYIDSETPIPG--EMGTLMLALAFCEAG 185	
132	VWITATSPKQGTGTGWMFYVAFSAGLPTIGALVATISMQVYNLTFEYETLWASLVVIG 191	
166	GVYAMISLRHVHTPGHMHNLTPREKFAELSRVAVTLTYTRNIFLSSIVAITWLSIFGA 245	
192	STIALGVKE-RRGRRLPVANPDVVKQTGGQFKLLRNDRRARFVYIHTINSIPYAA 250	

```

OY      246 VIMPMFVDELGETSEMLQWAAFEFTTIE-----NIFMIVAEKMGMMVIRNGCL 300
Db      251 VFPPSFEEFDL-----KKQLSMELLITVIYAANLPNPFPESFEDRCHMATVWGSI 305
QY      301 GMAASSLAIFYVMPOFGHNWMA-----MIPAL-CTFYVAFVPMAAFALEPKKH 352
Db      306 GGAATLALVLYPIPM-FGVQAGSMGNCVGITIAAGLFCGLSVPLSAIAVISDPPKH 364
QY      353 GAATSVNYLSGMSNFLPALAVLLPPEFTIGVIATATALLAFVLCAFLRVDPGS 412
Db      365 GAAMATYNLVGGAAVAPBLVAVFHPILGPTGLILMALIYLISGMMTLRLGTGPGRD 424
QY      413 SAPYTEKALNI 423
Db      425 GVPALEDAHI 435

RESULT 4
AY37671
ID AAY37671 standard; Protein: 414 AA.
AC AAY37671;
XX
XX 07-OCT-1999 (first entry)
DE Chlamydia trachomatis transport protein.
XX
XX Vaccine; eye disease; conventional trachoma; nongonococcal trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perithea-
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholin's; pneumonia; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
XX W099928475-A2.
XX
XX 10-JUN-1999.
XX
XX 27-NOV-1998; 98MO-IB01939.
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
XX (GENSET ) GENSET.
XX
XX Griffiths R;
XX
XX WPI; 1999-371125/31.
XX
XX Genome sequence of Chlamydia trachomatis
XX
XX Disclosure: Page 1298-1300; 1755bp; English.
XX
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
PS of Chlamydia trachomatis (see AY01425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nongonococcal trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perithea-
CC patitis, Bartholin's; pneumonia in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
XX Sequence 414 AA;
XX
XX Query Match | 7.9%; Score 174.5; DB 20; Length 414;
XX Best Local Similarity 22.8%; Pred. No. 7.5e-10;
XX Matches 98; Conservative 63; Mismatches 150; Indels 119; Gaps 22

```

QY 40 IKSLGFTPAEAS-FAFTLYGLAALISAVSGVAEITTPOKTMLIGFVLMCFVHVLVE 98
DB 18 IADLFEDFAQAGIIGSTLY-ITYGISKEVSGVMSQSNPRFYMAIGLITGTSNIF--F 73
QY 99 GLGQANYGLILFVGI-----KGLAEP-----LELYSFIYIIHNVSSENSSALGMYMAVY 150
DB 74 GL-STIPLFVLFKINCFQMGWMPPCARLITHWY-----SKSERGTWMSVW 120
QY 151 SVGIGVAGSYIPSTF--IPIMGEMGLMLALPFCAGVIAMISLRHVKTGPHHNLTP 207
DB 121 STSHNIGGALIPVLGVADYTGWGVMEIPGIIICITINGFILLRLR--DIPQSL-GLPA 177
QY 208 REKFELSRATVLLTYNNIFLSSIVRLIN-----LSLGFPAVIMPM 251
DB 178 IEKFEKEDAPHEETTDIEEERELSTKEITFVLSNKMWLFSPASFYIYVVM 237
QY 252 FVDELGFTSEMLOYMAFF-----FTTFSNI-----FMGIYAE----- 286
DB 238 AVND-----WSALYLETKDYSTVKANLCVSLFEIGGLFGMLLAGLSDPTISK 286
QY 287 KMGWNRVIRMFGLGMAASSLAFFYMPQYFGHNVMAMIPALGTFVAAPVMAAV-PP 345
DB 287 KRGPNNVY--FSLGLLVSIILGIMGTROYF--VMMIDGTFLEIGFLL--FGPQMMIGLA 339
QY 346 ALEPKHGAATSVYVLSGMSNPLAPALAVVLLPWFSTIGVIATL-----Y 394
DB 340 AAELSHKRAA---GTASGFTG-----WFAYFGAFAAGYPLGVAQDMGWHGF 383
QY 395 LLAFLVCAFI 404
DB 384 FVALLACALI 393

RESULT 5
AAB18235
ID AAB18235 standard; Protein: 457 AA.
XX AAB18235:
XX 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related protein seq ID NO:92.
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX Plasmodium falciparum.
OS MO200025728-A2.
XX 11-MAY-2000.
PD 05-NOV-1999; 99WO-US26796.
PF 05-NOV-1998; 98US-0107131.
XX (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI: 2000-365347/31.
DR
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure: Page 220-221; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 457 AA;
Query Match 7.6%; Score 168.5; DB 21; Length 457;
Best Local Similarity 19.8%; Pred. No. 3.8e-09;
Matches 90; Conservative 78; Mismatches 154; Indels 133; Gaps 18;
QY 35 FLSHYIKSLGTPA---EASFAFTLYGLAALISAVSGVAEITTPOKTMLIGFVLMCFV 91
DB 38 YVISYMKIIGCSVDKRYKSSWIVLTLFQCFEFGFGLINQNGPQISVLGGMLKGLG 97
QY 92 HVL--FLVFGAGQANYGILFYGIR-----GLAYPLFYFIYVYIIHNVSSENSSALG 144
DB 98 ILSYFTVF-----NFYLFMTYGLIGCGIAYPIPL-----SVAVKKHYDKG 143
QY 145 WYMAVYSGVIGVA-----GSYIPSE-----TIPMGEMGLML 177
DB 144 VISGIIIFGRSLVFIICPLQNYINKNYMPDMPLEIENDEKTFSLDLINKVPILEI 203
QY 178 ALAFCFAGGVIAMISLRHVKTGPHHNLTPREKFAELSRATVLLTYNNIF-----LSS 231
DB 204 YEGICFPA--IIQFLGSYLIADSGD---TSKDFAMAYNDRNKKVLYFEKKINPKNGLSN 257
QY 232 IVRIINTLSLGFVAVIMPMFVDELGFTSEMLOYMAFFT----- 273
DB 238 SLRFLSNTSNFSREV-----NNTFLNREFLLIWLMEFNMOAISYTOYFWKIFGKNY 310
QY 274 -----TIFSNIFWGIVAEKKMGW-----MRYIRMFGLGMAASSLAF 309
DB 311 LSIIDRSLSLGVSYLENIFGRIFWGLISDPTSFKTLIIMSLMSF--LTITLTMSCF 368
QY 310 YVMPQYFGHNVMAMIPALAGTFVAAPVMAAVPALE---PKHGAATSVYVLSAG 364
DB 369 YGIIITY---SIWCLIEFCHAGTF-----AIFPSITAHFTGKNGDPVGLFTARA 417
QY 365 MSNPLAPALAVVLLPWFSTIGVIAATAYLLAAV 399
DB 418 FSSIIMAIISAVL---NNIGINAMCAIVLSSEV 449

RESULT 6
AAB47287
ID AAB47287 standard; Protein: 394 AA.
XX AAB47287:
XX 31-JAN-2002 (first entry).
DE Enterococcus faecalis polypeptide NORA.
XX MDR; efflux pump; multidrug resistance; antibacterial; drug target.
XX Enterococcus faecalis.
OS
XX

Db 316 VTTYLKSHAGNRGAINGLNSAFTSFGNLLGPMAGYMPDLNHLFPYYISAIILIGTGFL 375
 QY 392 ALYL 395
 Db 376 SLFL 379

RESULT 8
 AAY02610
 ID AAY02610 standard; Protein: 416 AA.
 AC AAY02610;
 DT 21-JUL-1999 (first entry)
 DE Amino acid sequence of a putative efflux protein.
 XX
 XX Putative efflux protein; Pseudomonas mendocina KR-1; resistance;
 KM solvent; antibiotic; p-hydroxybenzoic acid.
 XX
 OS Pseudomonas mendocina.
 XX
 PN WC9923224-A1.
 XX
 PD 14-MAY-1999.
 XX
 PF 30-OCT-1998; 98WO-0523266.
 XX
 PR 31-OCT-1997; 97US-0961738.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Chen KK;
 XX
 DR WPI: 1999-313345/26.
 DR N-PSDB: AAX59540.
 XX
 PT New putative efflux protein gene increases resistance to solvents
 and antibiotics.
 PS
 PS Claim 4; Page 32-33; 28pp; English.

CC The present sequence represents a putative efflux protein obtained
 from Pseudomonas mendocina KR-1. The protein increases resistance to
 solvents and antibiotics, and is used particularly for constructing
 CC strains that produce p-hydroxybenzoic acid. Overexpressing the efflux
 CC system or its expression from a plasmid vector increases resistance of
 CC bacteria to a variety of toxic substances, while inactivating an efflux
 CC system increases sensitivity to antibiotics and toxic substances.
 CC
 XX
 SQ Sequence 416 AA:

Query Match 7.2%; Score 158.5; DB 20; Length 416;
 Best Local Similarity 22.7%; Pred. No. 3.9e-08;
 Matches 88; Conservative 56; Mismatches 164; Indels 77; Gaps 21;

QY 20 IALAIVFMG-----DGFELAFLSHYIKSLGFTPAEASFATLVG-----LAAL 63
 Db 66 VAVVVLTTGLVAATYGGIYLAQIVPDM---QASLGFSGSYVGTITASAQLGFLCAML 122
 QY 64 SAWSSGVAELTPQ---KIMLIGFVLMCVFHYFLVFGAGQANTGLLFGIRGLATP 120
 Db 123 AAW-----LPPKVGGRMTAASGACALALLPLLSNTFVIGVLTLL--LAGTAAT 172
 QY 121 LFLVSPFVITIHNVRSNSSSALGMYAVYSGIGVAGSYIPSTTIDMGEMGLMLALA 180
 Db 173 VFW-PWDVIAKRVAYKRGIAKGLVSSGTSYGAINSLVPIYA--PQGEKRSVMWVG 229
 QY 181 -FCFAGGVIMISLRHKVTPGHHNLTLPREFALSRVLTLLYNNRIFLSSIVR----- 234
 Db 230 LRTIAMVIVLVVL---KRGLLGOALPSAPATATDEAS-----SSGLSGLRPPVLL 280
 QY 235 IINTLSLFGRAVIMPMKVD-----ELGFTTSEMLQVMAAFETTTFISNIFWGIABKMG 289

Db 281 IWSMNFILGFAITPEFQYTLSSYLRTLEGFDVQYTAQVWAIGFVGWAGLAVGMLSDRTG 340
 QY 290 WMRVIRW--FGCLMAASSLAIFYMPOYFGHNWMAIMIPIALCTEFAPVMAAVPAL 347
 Db 341 -LRFAMLYVYGCYVTA--LIFVYQPS--GH--W-----PLVAVLSTAFYPIFGILPAY 389
 QY 348 EPR--HKGAISVYNL-----SAGM 365
 Db 390 VSKLASSAMAVSTFGIANVMOGSGM 416

RESULT 9
 AAB98023
 ID AAB98023 standard; Protein: 416 AA.
 AC AAB98023;
 DT 16-AUG-2001 (first entry)
 DE Pseudomonas mendocina KR-1 putative efflux 4.5X protein SEQ ID NO:2.
 XX
 XX Pseudomonas mendocina KR-1; efflux; 4.5X; RFLP marker; genetic mapping;
 KM restriction fragment length polymorphism marker; physical mapping;
 XX solvent tolerance.
 XX
 OS Pseudomonas mendocina.
 XX
 PN US6225089-B1.
 XX
 PD 01-MAY-2001.
 XX
 PF 15-JUN-1999; 99US-0333208.
 XX
 PR 30-OCT-1998; 98US-0183270.
 PR 31-OCT-1997; 97US-0961738.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Chen KK;
 XX
 DR WPI: 2001-342396/36.
 DR N-PSDB: AAB21716.
 XX
 PT Novel isolated Pseudomonas efflux polynucleotide, useful for probing an
 organism's efflux system to gain understanding of the mechanism of
 PT solvent tolerance and as restriction fragment length polymorphism
 marker
 PS
 PS Claim 1; Column 23-26; 17pp; English.

CC The present sequence represents a putative efflux protein, designated
 CC 4.5X, isolated from Pseudomonas mendocina KR-1. The putative efflux 4.5X
 CC gene (I) is located immediately downstream of the poba gene in
 CC Pseudomonas mendocina KR-1. The present invention also describes: (1) a
 CC chimeric gene (II) comprising (1) operably linked to at least one
 CC suitable regulatory sequence; (2) a transformed host cell (III)
 CC comprising (II); (3) isolation of (I); and (4) a complement of (I). (II)
 CC is useful for altering the level of expression of Pseudomonas efflux
 CC protein in a host cell by transforming the cell with (II), and growing
 CC the transformed host cell produced under conditions that are suitable
 CC for expression of (II), resulting in production of increased levels of
 CC Pseudomonas efflux proteins in the transformed host cell relative to
 CC expression levels of an untransformed host cell. (I) is useful for
 CC isolating cDNAs and genes encoding a homologous putative efflux 4.5X
 CC gene from the same or other bacterium species. (I) is useful as a
 CC restriction fragment length polymorphism (RFLP) marker, and for carrying
 CC out nucleic acid amplification-based methods of genetic and physical
 CC mapping. (I) is useful for probing an organism's efflux system to gain
 CC an understanding of the mechanism of solvent tolerance.

Sequence 416 AA:

XX Telford J, Maslmani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tellein H;
DR WPI: 2002-352536/38.
DR N-PSDB: ABN69435.
XX
PT New Streptococcus protein for the treatment or prevention of infection
or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 1: Page 3841: 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I), nucleic acids encoding (II), ABN66044-ABN71516 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX
SQ Sequence 397 AA:

Query Match 7.0%; Score 155; DB 23; Length 397;
Best Local Similarity 23.6%; Pred. No. 8.6e-08;
Matches 100; Conservative 76; Mismatches 161; Indels 86; Gaps 24;

QY 8 WGLPLHLINGYIAIAVFMGDFELA--FLSHYIKSLGFTPAE----ASFAFTLYGLA 60
DB 11 WLG-----NFTGASFSLVMPFMALYVENLG--TPRELVEYAGLAVAAYALA 56
QY 61 AALSAMWSGVVAELITPOKTMU--IGFVLMCVFHLFLVFGIG--QANGGLILFPGIGLA 118
DB 57 SALFAPRWGLADRYGKRPMLLRASFVW-----TFMGGALITPNVFWLLILLRLGVS 110
QY 119 YPLVLSFVIIVIHNVSESSSALGWYMAVSVGIGV--AGSYIPSTPIPMGEMTLM 177
DB 111 AG-YVVPATALLASQAARKESGYALG-----TLATGTGALLIGPLGLGLAEL--LGI 161
QY 178 ALAFCEPAGVYIAMISLR---HVKTPEGHMLTPREKFAELSRAVTLLYTRN-----NIFL 229
DB 162 ROVELLGVLTFLFCSLMTAVYVK-----EEFKPYRREMIPTKVIILQOVKSPOIMLGLFV 216
QY 230 SSIVRIINTSLFGEFAYIMPMFVDELG-----FTTSEMLOWMAAFETTISNIMWGI 283
DB 217 TSMITQISAOVS---APIL-SLYIRHLGOTHLNMTFTSG---LVVASGFSLSSTSLGK 269
QY 284 VAEKMGMYRIYRFGCLGMAASSLAF--YVMPDYFGHNYMMAMIPALGTFFVAFPV-MA 341
DB 270 LGRBFNGHRL-----LALCYSFIMYPSALSALQTSFGLGVRFRFYGVGGVGLMPSIN 322
QY 342 AVFPALPEPKKGAIAISYVN--LSAGMSNFLAPALAVLLPMPSTIGVIAVTALY----LL 396
DB 323 SLTLKTLTPKGIISRVFAYNOMFSLNGQVIGPFIG-----SNVAVVLGYSRVFVYVSLI 375
QY 397 AFW 399
DB 376 VFV 378

XX AAE05852;
AC 24-SEP-2001 (first entry)
DT XX
DE Pseudomonas stutzeri open reading frame-N (ORF-N) protein.
XX
KW Open reading frame-N; ORF-N: pyridine-2,6-bis (thiocarboxylate);
KW PDTC; environmental remediation; phytoremediation; bioaccumulation;
KW water purification; solution mining mobilisation; immobilisation;
KW detoxification; redox state modifier; metal ion reactivity;
KW carbon tetrachloride; metal.
XX
OS Pseudomonas stutzeri.
XX
XX WO200153309-A1.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001MO-US02386.
XX
PR 20-JAN-2000; 2000US-0177251.
XX
PA (IDAH-) IDAHO RES FOUND INC.
PA (PASZ/) PASZCZYNSKI A.
PA (SEBA/) SEBAT J L.
XX
PI Paszczynski A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;
XX
DR WPI: 2001-465361/50.
DR N-PSDB: AAD11172.
XX
PT New isolated nucleic acids, useful for producing enzymes required to
PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
PT reducing the amount of metal or carbon tetrachloride in a substrate,
PT e.g. soil or water -
XX
PS Claim 32: Page 100-101; 172pp; English.
XX
XX The present sequence is a Pseudomonas stutzeri open reading frame-N
XX (ORF-N) protein. The Pseudomonas stutzeri genome includes ORFs that
XX encode enzymes required for biosynthesis of pyridine-2,6-bis
XX (thiocarboxylate) (PDTC). The ORFs encoding PDTC are especially useful
XX in environmental remediation methods, e.g. phytoremediation,
XX bioaccumulation, water purification, waste water purification, solution
XX mining mobilisation, immobilisation, detoxification, redox state
XX modifier or modification of metal ion reactivity. In particular, the
XX ORFs are useful for degrading carbon tetrachloride and removing metals
XX from substrates, e.g. soil or water.
XX
SQ Sequence 392 AA:

Query Match 6.9%; Score 152.5; DB 22; Length 392;
Best Local Similarity 21.6%; Pred. No. 1.6e-07;
Matches 89; Conservative 68; Mismatches 182; Indels 73; Gaps 16;

QY 9 LGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASF--AFTLYGLAAALSAW 67
DB 22 MGDPMIMFY-----ALCITLG-----PHLVADLGISROOLGMLTASTGALALLSPWA 68
QY 68 SGVVAELITPOKTMUIGFVLMCVFHLFLVFGIGQANGGLILFPGIGLAFLFLXFT 127
DB 69 GALVORMGCT--RAGLI-----CNFLVGL---SFSIMAVLPGGGVLTALLCGTA 114
QY 128 -----VVIHNVSESSSALGWYMAVSVGIGVAGSYIPSTPIPMGEMGTL--W 176
DB 115 QSLANPATNOAIAHSVPAKAGVGLKQSGVQASALLAGVALPLVL--WNGRGALAAW 173
QY 177 LALAFCEPAGVYIAMISLRHVKTPEGHMLTPREKFAELSRAVTLLYTRNIFLSSI--VR 234
DB 174 VVALVMAALVYVWPAKKSAP-----SLPLRVNGRPVWMLSTILMAIQ 216
QY 235 IINTSLFGEFAYIMPMFVDELGFTTSEMLOWMAAFETTISNITWGIYAEKMGMYRI 234

```

Db      217  LCAGLALSFMTFLG-VYAAQIGVSVSTIGAMVSCFGAMGILSRVLLFTPLADLKDETLI 275
OY      295  RMEGCLGMAASSLAIFYMPOYFGHNWMMIPAIAGTGYAAVPMAAVFPALPEKHKA 354
Db      276  --LGVLFTLA-GLAALVMEANTQOHPMLGLGTGMLTVAASNAIAMSLLDGRFGA 332
OY      355  AISVYVLSAGMSNFLA---PATAVVL-----LPMFSTGVVAVTALYLL 396
Db      333  AHSAGMLSTVGFGGFAVGPFAFGMFLAHSEGFPAAMLSLIGLIVAGGLCLL 384

RESULT 13
ID      ABB48316 standard; Protein; 401 AA.
AC      ABB48316;
XX      05-FEB-2002 (first entry)
XX      Listeria monocytogenes protein #1020.
DE      Listeria monocytogenes protein #1020.
XX      Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX      vitamin B12; bacterial infection; disease.
OS      Listeria monocytogenes.
XX      MO200177335-A2.
XX      18-OCT-2001.
XX      11-APR-2001; 2001WO-FR01118.
XX      11-APR-2000; 2000FR-0004629.
XX      (INSP) INST PASTEUR.
XX      Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX      Dussurget O, Chetoui F, Nedjar H, Glaser P, Kunst F, Cossart P;
XX      Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
XX      Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX      Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L;
XX      Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX      Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX      Rose M, Voss H;
XX      WPI; 2002-010914/01.
XX      Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX      and prevention of Listeria and related bacterial infections, and
XX      related polypeptides
XX      Claim 6; SEQ ID No 1021; 192pp; French.
XX      The present invention relates to the genome sequence of Listeria
XX      monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX      it are useful for selecting probes and primers for detecting genes in L.
XX      monocytogenes and related organisms, and for studying genetic
XX      polymorphisms and other genomes. The present sequence is a protein
XX      encoded by the genome sequence of the present invention. Proteins
XX      expressed from the genome sequence are useful for raising specific
XX      antibodies. Identification of L. monocytogenes and related organisms, and
XX      for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX      B12. The genome sequence and proteins encoded by it are also useful for
XX      selecting compounds that regulate gene expression and cell replication
XX      and modulate L. monocytogenes-related diseases. In addition, the genome
XX      sequence and proteins encoded by it are useful in pharmaceutical and
XX      vaccine compositions for the treatment or prevention of infections by L.
XX      monocytogenes and related organisms.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.

```

```

SQ      Sequence      401 AA;
Query Match      6.6%; Score 147; DB 23; Length 401;
Best Local Similarity 23.6%; Pred. No. 6.2e-07;
Matches 45; Conservative 40; Mismatches 88; Indels 18; Gaps 4;

OY      224  NRIIFLSIVRIINTLSLGFVIMPMFVDELGFSTSEMLOVMAAFPTIFSNIFMG 283
Db      7    NKGFVITLINFVLYVLYLLMVIIVIAIOEELNALSGLGEPASGIYITGLARLYMK 66
OY      284  VAEKMGMRVIRMEGCLGMAASSLAIFYMPOYFGHNWMMIPAIAGT-----FVAA 336
Db      67  KLELFGRKKVLR-FGILFELITTMAYLMPT-LAIWEIIRFLNGFAVGTSTATNAIVTA 124
OY      337  FVPMMAVFPALPEKHKAISVYNLSAGMSNFLAPAIAVVLLPMFSTIGVAVTALYLL 396
Db      125  YIPNS-----RNGEGINYYGLSTSLAAIGPFIIGMLLSKTSFYTIITFIVVLL 175
OY      397  AFVLCAFIRVE 407
Db      176  TALCFYLPVK 186

```

```

RESULT 14
ID      ABP40749 standard; Protein; 393 AA.
AC      ABP40749;
XX      24-JUL-2002 (first entry)
XX      Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5594.
XX      Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX      antibacterial; gene therapy.
XX      Staphylococcus epidermidis.
OS      Staphylococcus epidermidis.
XX      US6380370-B1.
XX      30-APR-2002.
XX      13-AUG-1998; 98US-0134001.
XX      14-AUG-1997; 97US-055779P.
XX      08-NOV-1997; 97US-064964P.
XX      (GENO-) GENOME THERAPEUTICS CORP.
XX      Doucette-Stamm LA, Bush D;
XX      WPI; 2002-381255/41.
XX      N-PSDB; ABN93294.
XX      Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX      polypeptide, useful for diagnosing and treating bacterial infections -
XX      Disclosure; SEQ ID 5594; 267pp; English.
XX      ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX      frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX      given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX      antibacterial activity and can be used in gene therapy. The sequences
XX      can also be used in the diagnosis and treatment of bacterial infections,
XX      particularly S. epidermidis infections. The sequences can be used to
XX      screen for compounds able to interfere with the S. epidermidis life
XX      cycle or inhibit S. epidermidis infection.
XX      N.B. The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from the
XX      USPTO web site.
XX      Sequence      393 AA;

```



```

Db 282 IGYLSNFQMGRTMSADG---TAGEWLGWTFYWMWMISSPFGMLARISGRS 338
QY 294 IR-----WFCGLMASSLAFFYMPQYFCHNYW----- 321
Db 339 IREFIIGVLLVPAGVSTWPSIFG--GTAIVF---EONGESIWDGGAEEQLFGLHAL 392
QY 322 ----MAMIPALGTFVAFVPMVAVPALPEPKH-----KGAISVYNL 361
Db 393 PGQIMGITIMILGTFFTISADSASIVMCTMSQHGLEANKWVTAMGVATAIGLTL 452
QY 362 SAGMSNFLAPAIIVLL---PW--FSTIGVIA 389
Db 453 LSGDNALSNLQWVTIYAATPFLFVYIGLMFA 484

```

Search completed: March 13, 2003, 16:53:13
 Job time : 36.7483 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 11.5828 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNKQWLGLEPLHIMGYI.....RYEGPQSSAPYTERKALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1040.5	47.1	467	US-09-134-001C-3020	Sequence 3020, Ap
2	158.5	7.2	416	US-09-333-208-2	Sequence 2, Appl
3	158.5	7.2	416	US-09-333-254-2	Sequence 2, Appl
4	158.5	7.2	416	US-09-183-270-2	Sequence 2, Appl
5	145	6.6	393	US-09-134-001C-5594	Sequence 5594, Ap
6	136	6.2	492	US-09-134-001C-4847	Sequence 4847, Ap
7	132.5	6.0	423	US-09-134-001C-5210	Sequence 5210, Ap
8	126.5	5.7	473	US-09-134-001C-3564	Sequence 3564, Ap
9	125	5.7	477	US-09-134-001C-3487	Sequence 3487, Ap
10	124	5.6	479	US-09-134-001C-4435	Sequence 4435, Ap
11	123.5	5.6	515	US-08-063-552-4	Sequence 4, Appl
12	123.5	5.6	515	PCT-US93-05704-4	Sequence 4, Appl
13	123	5.6	418	US-09-030-267-5	Sequence 4, Appl
14	122	5.5	499	US-09-134-001C-5370	Sequence 5370, Ap
15	121.5	5.5	412	US-09-134-001C-3949	Sequence 3949, Ap
16	119.5	5.4	514	US-08-063-552-13	Sequence 13, Appl
17	119.5	5.4	514	PCT-US93-05704-13	Sequence 13, Appl
18	119.5	5.4	534	US-09-031-392-4	Sequence 13, Appl
19	119.5	5.4	534	US-09-299-549-4	Sequence 4, Appl
20	119.5	5.4	534	US-09-610-417-4	Sequence 4, Appl
21	117.5	5.3	413	US-09-134-001C-5503	Sequence 5503, Ap
22	117	5.3	668	US-09-134-001C-3430	Sequence 3430, Ap
23	114	5.2	177	US-09-134-001C-4482	Sequence 4482, Ap
24	114	5.2	400	US-09-134-001C-2912	Sequence 2912, Ap
25	113.5	5.1	518	US-09-134-001C-4744	Sequence 4744, Ap
26	112	5.1	831	US-08-677-734A-11	Sequence 11, Appl
27	112	5.1	831	US-09-097-053-11	Sequence 11, Appl

28	110	5.0	429	2	US-08-677-049-5	Sequence 5, Appl
29	110	5.0	516	1	US-08-356-340-4	Sequence 4, Appl
30	110	5.0	516	2	US-08-786-555-4	Sequence 4, Appl
31	108.5	4.9	521	1	US-08-063-552-2	Sequence 2, Appl
32	108.5	4.9	521	5	PCT-US93-05704-2	Sequence 2, Appl
33	108	4.9	462	2	US-08-898-976-2	Sequence 2, Appl
34	108	4.9	462	2	US-08-898-976-4	Sequence 4, Appl
35	107.5	4.9	662	2	US-09-134-001C-4074	Sequence 4, Ap
36	107.5	4.9	1040	4	US-09-134-001C-5365	Sequence 5365, Ap
37	107.5	4.9	408	4	US-09-134-001C-4130	Sequence 4130, Ap
38	106	4.8	367	4	US-09-134-001C-5557	Sequence 5557, Ap
39	106	4.8	466	4	US-09-134-001C-3355	Sequence 3355, Ap
40	106	4.8	513	4	US-09-097-889-15	Sequence 15, Appl
41	106	4.8	620	1	US-08-301-722A-2	Sequence 2, Appl
42	105.5	4.8	381	4	US-09-134-001C-3743	Sequence 3743, Ap
43	105.5	4.8	398	4	US-09-134-001C-4353	Sequence 4353, Ap
44	104.5	4.7	493	2	US-09-031-392-10	Sequence 10, Appl
45	104.5	4.7	493	4	US-09-299-549-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-3020

Sequence 3020, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

```

OY      369 LAPAIWVLEWFEITGVIATATLYLLAFVLCAFIRVEOP 409
          : | | : : | | | | | | | | | |
DB      415 MGPGLTFLFATLGIGVWVIYAGLYVAGAVLFEFIPVQP 455

```

RESULT 2

```

US-09-333-208--2
: Sequence 2, Application US/09333208A
: Patent No. 6225089
: GENERAL INFORMATION:
: APPLICANT: Chen, Kevin K.
: TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene from
: TITLE OF INVENTION: Pseudomonas Mendocina
: FILE REFERENCE: CL-1160-C
: CURRENT APPLICATION NUMBER: US/09/333,208A
: CURRENT FILING DATE: 1999-06-15
: EARLIER APPLICATION NUMBER: 09/133,270
: EARLIER FILING DATE: 1998-10-30
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 2
: LENGTH: 416
: TYPE: PRF
: ORGANISM: Pseudomonas mendocina KR-1
US-09-333-208--2

```

Query Match

Best Local Similarity 22.7%; Pred. No. 9.8e-08;
Matches 88; Conservative 58; Mismatches 164; Indels 77; Gaps 21

QY 20 IAINVFMG-----DGFPLATLSHYISLGFTPADSPAPFLVY-----LMAAL 63
 Db 66 VAVVYLLTGLVAATYGGBOITYLFAQLVPM---QASLGFSGFYSTTTASAOGLGLLAML 122
 QY 64 SAWVSGVVAEITTPQ---KTMLGIFVLMCFHVHFLFVGLGQANYGILLFFYGRGLAYP 120
 Db 123 AVW-----LTPKVGSGRMIAASGAVALALLILPLSSNFFVIGVLLT--LAGTAAT 177
 QY 121 LFLVSLFYIITHNRYSEMSSSALGMYAAVYVSGGVAGSYTPSTTPIRMGMGLMLALA 180
 Db 173 VEV--PMVDVIARVVAAYRYRGLAMGLVSSGTSYGAIVNSLAPRYA--PQGEWRSVWVWG 225
 QY 181 -FCEAGGVIAMISLRHYVTPGHMNLPRREKFAELSTRVLTLLYRNMLFSSIYR----- 234
 Db 230 LTLTAMVYLVLYVL---KRGGLLQALPASPATATTTDEAS-----SSGLSGLTRPVLL 280
 QY 235 IINTLSLFGFAVIMPMFVD-----ELGFTTSEMLQVAAAFEFPTTISNIFWGVIAEKMG 285
 Db 281 IMSNMFLIGFATPPQGYLTSSLYLRETELGEFDVQYQAWAVYIGFGMFGAGLAVGLSLDRTG 340
 QY 290 WMRVYIRW--FECCLGMAASSLAFYTMPOFGHINYMMAMIPALATGTEVAALVPMAAVFPAL 347
 Db 341 -LRFAMYLTVYCQVYTA--LIEFVQPS--GH-W----PLVAALFSTAFPIRGLIPAY 385
 QY 348 EPK---HKGAISVYNL-----SAGM 365
 Db 390 VSKLASSAMAAVSIFFGLANVMQSGGGM 416

RESULT 3

```

US-09-333-254-2
: Sequence 2, Application US/09333254A
: Patent No. 6235882
:
: GENERAL INFORMATION:
:
: APPLICANT: Chen, Kevin K.
:
: TITLE OF INVENTION: A Gene Encoding a Putative Efflux Protein for Solvents/
:
: TITLE OF INVENTION: Antibiotics in Pseudomonas Mendocina
:
: FILE REFERENCE: CL-1160-D
:
: CURRENT APPLICATION NUMBER: US/09/333,254A
:
: CURRENT FILING DATE: 1999-06-15
:
: EARLIER APPLICATION NUMBER: 08/961,738
:
: EARLIER FILING DATE: 1997-10-31
:
: NUMBER OF SEQ ID NOS: 6

```

```

; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 416
; TYPE: prt
; ORGANISM: Pseudomonas mendocina KR-1
US-09-333-254-2

```

Query Match

Best Local Similarity 22.7%; Pred.No. 9.8e-08;
Matches 88; Conservative 58; Mismatches 164; Indels 77; Gaps 21

OY		20	I A I V F M T G-----D G E L A F L S H Y I S I G F T P P A E S A F P L V G-----L A A A L	63
D b		66	V A V V L L T G L V A A T A G F G C I Y L F A Q L V P D M---Q A S I G R G F S Y C T I T A S A Q L G L L C A M L	122
OY		64	S A W S G V A A E I T P O---K T M L I G F V L M C H N H U F L V P G L G A N N G L L L F Y G I R G L A Y P	120
D b		123	A W V-----L T P R V G G R M A A S G A V C A L A L L I P L S N T F E I V G L L L---L A G T A A T	177
OY		121	L E L S E I V I L I H N R S E N S S A L G M A Y A V S V G I G V A G S Y P S P T I R I M G E M T L A L A	180
D b		173	V F V--P M D V I A R V A Y A R R G L A M K L B V S S G T Y G V A I N S L V P I Y A--P O G E M R S V M W Y V G	225
OY		181	-F C P A G G V I A M I S L R H V K T P G H M H N L P R E K F A E L S R A V T L L Y T N R N I F S S I V R-----	234
D b		230	L L T L A M F L V L V L V L---K R G I L L G O A L P S A P T A T T D E A S-----S G L S G L R I P A V L L	280
OY		235	I I N T L S L E G F A V I M P M M V P D-----E L G F T T S E M L O Y W A A F F F T I R S I N I F M G I A E K M G	289
D b		281	I W S N F L I G F A T P P O Y I L S Y L T T E L I G F D V Q Y A Q W A V A I G F G M R A G L A V H G L S D R T G	340
OY		290	M W R Y I R W--F S C L G M A A S S L A F Y Y M P O Y E G H N Y M M A M I P A L A I G T F V A A F V P M A A V P A L	347
D b		341	-L R F A M Y L V Y C G C V Y T A A--L I F V V O P S--G H -N----P L V A A L F S T A F P I G L I P A Y	389
OY		348	E P K---H K C A I S Y N L-----S A G M	365
D b		390	V S K L A S S A M A V S I F G L A N Y M Q G S G G M	416

RESULT 4

```

US-09-183-270-2
:
: Sequence 2, Application US/09183270B
: Patent No. 6410265
:
: GENERAL INFORMATION:
:
: APPLICANT: Chen, Kevin K.
:
: TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene
:
: TITLE OF INVENTION: from Pseudomonas Mendocina
:
: FILE REFERENCE: CL-1160-A
:
: CURRENT APPLICATION NUMBER: US/09/183,270B
:
: CURRENT FILING DATE: 1998-10-30
:
: EARLIER APPLICATION NUMBER: 08/961,738
:
: EARLIER FILING DATE: 1997-10-31
:
: NUMBER OF SEQ ID NOS: 6
:
: SOFTWARE: Microsoft Office 97
:
: SEQ ID NO 2
:
: LENGTH: 416
:
: TYPE: PRT
:
: ORGANISM: Pseudomonas mendocina KR-1
:
US-09-183-270-2

```

Query Match

Best Local Similarity 22.7%; Pred. No. 9.8e-08;
Matches 88; Conservative 58; Mismatches 164; Indels 77; Gaps 21

[illegible]

```

QY 121 LELSEFVITIIHNHRSSESSALGCMWAVYVSGVIGVAGSVIPSEPTIPIEMEMGIMTLA 180
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 VEV-PMVDVIARVAVAYRKLGLAMKLVSSGSIYGAISLLPPIYA--PQEMRSVMWVG 229
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 -FCFAGGVAMI SLRHVKTGPHMNLPRKFEALSHAVTLLEYTRNIFLSSIVR----- 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 LLTFLAMTVLYLVVL---KRGLLGOALPSAPTATIDEAS-----SSGLIIRPVLL 280
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 IINFLSEGFVIVPMNFVD-----ELGFTTSEMLQWMAAFPTTITSINFMGIVAEKMG 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 281 IWSNMFLEGFETPEFPQYLLSYLTETELGFDVQYTAQWAAVIGFWMGAGLAVGMLSDRTG 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 WMRYIRW--FCGLMSASLAFYMPQYFGHNWMMAMIIPAIALCTEVAALFVMAAVALPAL 347
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 -LRFAMTLVYGCVTAA--LIFVQPS--GH--W-----PLVAVLFTSTARYPIFGLIPAY 389
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 348 EPK---HKGAISVYNL-----SAGM 365
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 VSKLASSAMAVSTFGIIANWQGGGGM 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
US-09-134-001C-5594

```

: Sequence 5594, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5594
: LENGTH: 393
: TYPE: PRN
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-5594

```

Query Match	6.68;	Score 145;	DB 4;	Length 393;
Best Local Similarity	19.78;	Pred. No. 2e-06;		
Matches	81;	Conservative	76;	Mismatches 190;
			Indels	64;
			Gaps	19;

```

QY 10 GLPL--HLIMGVIAYAVEMTGDGFELAFSLSHYIKSLGFTPAESFAFTLYGLAALASAW 67
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
Db 4 GVPMMKOLFLEYFNFLFLFGLIGLIVIPVLPYTLKDLGLSGDGMVAAPFALSOMITSP 63
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
QY 68 SGVVAEITTPOKTMLIGFVLMCVFHVLPVLFVGLQANVGLLLFVGRILAAPLFLYSFI 127
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
Db 64 GGTATDKLGKLLIICIGLVE---FAVSEFMFAQG-SPTLLIISRVLGGESAGMWPVGT 119
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
QY 128 VVIHNVSSENSSSALGWTMAVYSVGIGVAGSYISFTPIPIGEMGTLLALAFCCAG-- 185
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
Db 120 GMIDISPGADKAKNFGMSAIIINSGF-ILG---PGFG---GFLAEISHRLPEFYAATL 171
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
QY 186 GVIAIIS---LRH---VTPRGHMHLTPREKFAELSRATLLTYTNINIFLSIV-RIINT 238
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 172 GVAAITMSVLLIHNPOKATTDGFHOYQ-ELFTKI-----NMKVETTPVILTLVIA 221
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 239 LSLFEGFAYIMPMPFDELGETTSE-WLQVWAFAFFTIPSINFVGIAGVAKMGMRVIRNF 297
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
Db 222 FGLSAPFELLFSLYTDKRWYTPKDISIALIIGGVGALQVFFEDFKFMYMSELNITAM- 280
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
QY 298 GCLGMAASSLAFYMPQYFGHNYMMAMIIPAALGTFVAAFVPAAPALP-----EPK 350
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
Db 281 ---SLYSALIVLML--VLANGYWTIMITISFYV-----FIGDMIRPALTNFYSNIACK 329
   | : - - - - - | : - - - - - | : | : | : | : | : | : | : | : | : | :
QY 351 HKGAALISYNNISAGMSNELAPALAVL-----LPMESFT-----GVVIAT 390
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

```

Db          330  RQ9RAGGAGGAGTSGMGNFICGLVAGALFDVNLFFPLVMAIVASLSGIIIF 380
          RESULT 6
          US-09-134-001C-4847
          ; Sequence 4847, Application US/09134001C
          ; Patent No. 6380370
          ;
          GENERAL INFORMATION:
          ;
          APPLICANT: Lynn Doucette-Stamm et al
          ;
          TITLE OF INVENTION: NOCETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
          ;
          TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
          ;
          FILE REFERENCE: GTC-007
          ;
          CURRENT APPLICATION NUMBER: US/09/134,001C
          ;
          CURRENT FILING DATE: 1998-08-13
          ;
          PRIOR APPLICATION NUMBER: US 60/064,964
          ;
          PRIOR FILING DATE: 1997-11-08
          ;
          PRIOR APPLICATION NUMBER: US 60/055,779
          ;
          PRIOR FILING DATE: 1997-08-14
          ;
          NUMBER OF SEQ ID NOS: 5674
          ;
          SEQ ID NO 4847
          ;
          LENGTH: 492
          ;
          TYPE: PRF
          ;
          ORGANISM: Staphylococcus epidermidis
          ;
          US-09-134-001C-4847

```

Query Match	6.2%;	Score 136;	DB 4;	Length 492;
Best Local Similarity	20.2%;	Pred. NO. 2.2e-05;		
Matches 96;	Conservative 82;	Mismatches 175;	Indels 122;	Gaps 22

```

0Y      1 MSRRNKKQGLPLHLIMGYIAIAVEMTGDGEL---ALSHYIKS---LCGTPAEASFA 53
Db      58 INENTSOHL-----ITGFMALVNGVMIPLITATLMDKVKTRPLYLG-----AMGS 100
QY      54 FTYGLGAAA-----LSAMVSGVVAEITTPQKTMILIGVLMCVFHYELVFLVGLGQAN 104
Db      101 FLIGSIYAIAIPNFGVLTIAIAVIOAVGAGIIMP-----LMOFTLLFTLFSKEHRGFMGLAG 156
QY      105 Y-----GILL-----LEYGIGLAPPLFLYSPFIYIIHNHVSSENS----- 139
Db      157 LVIOFAPIAGTTFGLLIIDNASMRVPFLIVIGIALVTFIFGVVASISSTNTTKETKLDKKS 216
QY      140 --SSALGMYMAVYVSGVIGASYSIPSEFTPIPMGEMGTLMALAFECFAGVIAMISLRHVK 197
Db      217 VIYSTLGGGLMLXYA--FSSAGNL--GFSNPITY--LCSLISLIDL-----IGIFVKRQIT 264
QY      198 TPGRMHNITPREKRAELSRAYTLTYTNRNIFLSSIVRIINLISLGFNAVIMPMHVEDELG 257
Db      265 ISNPLNKLK-----IFNNKIFCFSTTSMIMIMSMVGPLLLIPLYVQNMIG 310
QY      258 FTT--SEMIOVAAAFPEFTTIFSNIFEMGIVAECCKMGMRVIRWFGCGCJMAASSLAFYVMPQYF 316
Db      311 LSAALLSLGVIMPGAIIINIMS-VFTGRKYDKYG--PRPLIFGFIILLICTPLCLCKLKYDT 368
QY      317 GHNTYMMAMIPIAIALGTFAAFAVPM--AAVPEALEPKHKGAISVYNLSAGNSNFLAPAI 373
Db      369 SYTY---LIVYIAIRMFPAVSLIMPIINTGIALKTEDTISHGTALIMNNGRVAGSIGTAL 425
QY      374 AVVLL-----PMFST-----FGVYAV---TALYLAFLVCAFIR 405
Db      426 MYTFPISIAQWVYSSSEHASKEMIOROSVAGVDVSEFLAVTYEVILATVALEIK 480

```

RESULT 7
 US-09-134-001C-5210
 : Sequence 5210, Application US/09134001C
 : Patent No. 6380370
 : GENERAL INFORMATION:
 :
 : APPLICANT: Lynn Doucette-Stamm et al
 :
 : TITLE OF INVENTION: NOCETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: GTC-007
 : CURRENT APPLICATION NUMBER: US/09/134,001C
 : CURRENT FILING DATE: 1998-08-13

```

: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5210
: LENGTH: 423
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5210

Query Match
Best Local Similarity 22.7%; Score 132.5; DB 4; Length 423;
Matches 90; Conservative 60; Mismatches 123; Indels 124; Gaps 23;

QY 90 VFHVLFLVFG-----LGOAN-----YGLILFLYIGLAVPLFLYS 125
D 67 VFECVFLIFSPIGLKGRKNDKREFNTISMFAFLFSAGMGIGLVEY---GAAPMAHFA 123
QY 126 -----FLVLIHNVRENSSSALGW---YMAVSVGIGVAGSY-----IPSFT 165
D 124 APPLADETTKAYTESLR-----STFFHMGFHAMAIYGV-VALLALAYSQFRKGEPLISRT 178
QY 166 I-PIHGE-----KOTMLALA-PCFAGGVYAMISLRHVKTRGHMNL--TPREKFAE--L 214
D 179 LRPLLGKVEGPIGTLDIVSVFATLVGAVVSLGMLQINGGLHLYFGVPPNTEVGGII 238
QY 215 SRAVTLTY-----TNRNIFLSIYRIINTLSLFGFAVIMPMFVDELGFT 259
D 239 IYVVTILFISANSGLSKGYOYLSNLIIGCTILMIY-TLIVGPTVILNNM----- 289
QY 260 TSEMLOVMAAFETTTFSNIFWGIVAEKMGMYIRWEGCLGAASSLAFYMPQYEGHN 319
D 290 TSTGSLNSLSEFLNFSPTALNG---OKRDM-----STWTLTYW-----G 327
QY 320 YMMAMITAILGTFVA-----AFVPMAAVPALEPKHKAISVYNSAGM--SNF 368
D 328 WMLWSMSEFV--GVFIARVSKGRSIREFISGVLVLPALVSFTWESVFGVIGLACKDSSL 385
QY 369 LAPAIAVLLPWFESTI--GVFIATYATLYLAFVLCAF 403
D 386 FKMSPEQQLGCVFNHPIGLIVSLIALLIASLICYF 422

RESULT 8
US-09-134-001C-3564
: Sequence 3564, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3564
: LENGTH: 473
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3564

Query Match
Best Local Similarity 5.7%; Score 126.5; DB 4; Length 473;
Matches 100; Conservative 73; Mismatches 169; Indels 151; Gaps 26;

QY 1 MSRNKQWLG-LPLHLIMGYIAIVPMGDFELA--FLSHYKSLGFTPAEASF---A 53
D 30 LSVQRKLMKNFQAFVFEVYVYAMYLIRNNFKAAPLLK---EETGLTTLLEIGYIGLA 86
```

```

QY 54 FTL-YGLA-AALSAMVSG-----VVAEITPQKMLIGFVLCVHVLVEGLG 101
D 87 FSIITYGLGKTLIGYFVGGRNTRKRIISFLILSAT---VIMGFVL-----SYFG-- 133
QY 102 QANTGLILFYGIRGLAVPLEYFSIYVLIHNVSENSSA-----LGMVMA 148
D 134 -SYMGILLIVLMGNG-----IFQSVGPPASYSSTISRMARTRKRGYLGFWNT 179
QY 149 VYSGVIGVAGSYRSPITPIHMGEMTLMALAFCFAGGVYAMISLRHV----- 196
D 180 SHNIGALMAGV-----ALMGANTF-FRGNVGMFTFSPVIALIIGIVLFI 225
QY 197 -----KTRGHMHN-----LTPREKFAE--LSRAVTLTYNRIPLSSI 232
D 226 GKDDPEELGNWRABEIEEPRIDQENIDISOGMTKDKFKYILGNPVIWILCISNVF-YI 284
QY 233 VRIINTLSLGFAYVIMPMFVDELGFTTSEMLOVMAAFETTTFSNIFWGIVAEKMGMR 292
D 285 VRI-----GIDMMAELVYSEHLHFNKGDVNTIFFEIGALVALSLMGYISDLKGR 337
QY 293 YIRWFGCLGAASSLAFYMPQYEGHNYMMAMIPALAGTFVAAFVMAAVPALE---P 349
D 338 AIVAGCMEMITTFVLEFTN---ATSVYVNISLFLGALI--FGQLIGVSLTGFPV 391
QY 350 KHKGAISVYNSAGMSNF-----APAIAYVLLPWFESTIGV-----VIAYT 391
D 392 KN--AIVSANGMGTSFAVIFGDSMAKVGLAALDPTRNGINIGYTLGSDTVFIVFYV 448
QY 392 ALYLAFLVLCAFI 404
D 449 ALF-LGMILLAIY 460

RESULT 9
US-09-134-001C-3487
: Sequence 3487, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3487
: LENGTH: 477
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3487

Query Match
Best Local Similarity 22.2%; Score 125; DB 4; Length 477;
Matches 107; Conservative 74; Mismatches 172; Indels 128; Gaps 26;

QY 1 MSRNKQWLG-LPLHLIMG-----YI-AIAVPMGD-----GEELA 34
D 50 IETTAQWMLITGFMVINGIMIPLTAFLMDKYSFRNLTYIFSMALFLGSIIVAASPTTIL 109
QY 35 FLSHYIKSLG---FTPAESFAFTLY-----GLAALSANVSGVAIITPQKMLIGFV 86
D 110 MISRIIDAIAGLILPLMOFTVFLPAEQRGFAMGL---AGVYV-----QSPALIGPT 160
QY 87 LMCVFHVLFLVFGIQAANYGLILFLYIGIRGLAVPLFYSEIYVLIHNVSENSSS---A 142
D 161 LTGLFVDLF-----SMRMPFLVSAIAVAFLIGFFV-----ENNTKTKDIY 203
QY 143 LGMVAVYSV-GIGV---AGSYIPSPITPIHMGEMTLMALAFCFAGGVYAMISLRHVKT 198
```

```

Db      204 LDKISVYSTFGGLILFAFSSVSFTGTSPLVIVTEVGLIA-----IIIIFTQQLK 257
Oy      199 PGHMHLTPREKFAELSRVAVTLTYTRNIFLSSIVRIINTLSLFGFAVIMPMFVDELGF 258
Db      258 KHPLLNMR-----VEFNKRVTLSSVSSMLVYITWVSPALLIPYI--QTGL 301
Oy      259 TTESEMLQVAAFEFTTII--FSNIFMGIVAEKKMG--NMRVIRMEFCCLGMAASSLAFYMPQY 315
Db      302 GGSALLSGVVLPGAVINGLTWYTKIFDKHGIRKVLVLPF-----ILISMFLYSFL 357
Oy      316 FGHNTWMMAN-----IPALALGTFVAAFEVMAAV--FPALER--KH-----KG 353
Db      358 TGTPEYFVILYVYTRIMIALGLLV---MPLNTVGLNALSDDVSHGTAIMNSLRILAGMG 414
Oy      354 AAISSVYNISAGSNFLA--PATAVALLPWFSTI--GVVIAY---TALXLAFLVCATIRRE 407
Db      415 TVAVSYTILSVAKQYVASHSTMSKMLTQEAIVHGIDVAFITTVLIIIGLILAFIKKE 474
Oy      408 Q 408
Db      475 K 475

```

```

RESULT 10
US-09-134-001C-4435
; Sequence 4435, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4435
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4435

```

Query Match 5.6%; Score 124; DB 4; Length 479;
 Best Local Similarity 19.5%; Pred. No. 0.00034;
 Matches 86; Conservative 78; Mismatches 190; Indels 88; Gaps 20;

```

Oy      13 LHLIMGYIAIVMTGDFELAFSLHYIKSLGFTTPAASFAFTLGLAALASAWGQVYA 72
Db      59 LGIFICY--AGYLLKRNFSLAMP--LIEQGIKGLGIALSAVSIAVGFSGFVMTGYS 114
Oy      73 EITTPQKMLIGFVLMCVFHVLFVFGIGQANYGLILFY--GIRGLAYP---LFLY 124
Db      115 DRSNAMPFLGLVLAITINLLGIFPFTSSITIMEFLVGMWGQMGKMPSPSGVLYH 174
Oy      125 SPLVYIIHNVRSNSSLALGWTWAVSVGICVAGSYIIPSTPIPMGKMLALAFCA 184
Db      175 GFSV-----SERGSKTSIMWVA--HNVGCG-----LMAPIAIHWGISMTALYN 214
Oy      185 GG-----VIAMISLRHVKTPEGHMLNLPREK-----AELSRATLLY 222
Db      215 FGLKFEVGYIYPALLAIIIAIFSYILINDTPQSGCLPEIYQKNDYATSTKOTIETEL 274
Oy      223 TNRNIFLSSIVRIINTLSLFGFAVIMPMFVDELGFTTSEMLOY-----WA 268
Db      275 TTRKELFKYVLN---NKMWMAIAFTNIFVYVRGVLDMAPTYLSEKHPLDSAGNA 329
Oy      269 APFF--TTTFSNIFMGIVAEKK--GMNRYIRMGCGICMAASSLAFYMPQYFGHNTWMM 324
Db      330 YFLYEWAGIPGTLGCLDLEKFGRRGPAGFFFMGLGVITIFILYMLNRP--GH-AWLDN 386

```

```

Oy      325 IPALALGTFVAAFEVMAAVPALEPRKHGAISVYNISAGSNFLAPATAVALLPWFSTI 384
Db      387 LSLIGIGFLILYGV--MLIGIALDLYPKKAAGTAAGLT--GLFQYLGAVYMANIV-----L 439
Oy      385 GVVIAYTALYLAFLVCATIRV 406
Db      440 GFVYQHFQWH--IGFVLLTVYSI 460

```

```

RESULT 11
US-08-063-552-4
; Sequence 4, Application US/08063552
; Patent No. 5688936
; GENERAL INFORMATION:
; APPLICANT: Edwards, Robert H
; TITLE OF INVENTION: Vesicle Membrane Transport Proteins
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/063,552
; FILING DATE: 19930514
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Farber, Michael B
; REGISTRATION NUMBER: 32,612
; REFERENCE/DOCKET NUMBER: 9067-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-063-552-4

```

Query Match 5.6%; Score 123.5; DB 1; Length 515;
 Best Local Similarity 21.8%; Pred. No. 0.00042;
 Matches 89; Conservative 63; Mismatches 157; Indels 99; Gaps 23;

```

Oy      58 GLAALASAWY-----GVVAETITPQKTMILGFVLMCVFHVLFVFGIGQANYGLIIL 110
Db      133 GLIFASKATVQLTNFPIGLTRNIGYIPMPAGF---CIMFTSWFAF--SSSYAFLLI 188
Oy      111 FYGIRGLAYPLFLXSFVYIIHNVRSNSSLALGWTWAVS-----VGIGVAGSYI-- 161
Db      189 ARSLQIG-----SSCSYVAGGMGLASVYTTDDEKRGKPGIALGGLAMG 232
Oy      162 ----PSF--TPIPMGKMLTALAFCFAGVYIAMISLRHVKTPEGHMLTPREKFAEL 214
Db      233 VLVGPPGVSYLEFVGKTAFLVLAALVLDGALQLFVLQPSRVQESQKGT----- 285
Oy      215 SRAVTLTYTRNIFLSSIVRIINTLSLFGFAVIMPMFVDELGFTTSEMLOYAAFEFTT 274
Db      286 --LTTLLKDPYILIAAGSICFAN---MGIAMLEPALPIMMETMCSRKWQLGVAFLPAS 339
Oy      275 IF-----SNIFMGIVAEKKGMRYIRNF--GCLGMAASSLAFYMPQYFGHNTWMMAMP-- 326
Db      340 ISYLTGNTF--GILAKRMG-----RWLCALLGVYIVGISILCIP--FAKKYGLIAPNFG 391

```

QY 327 -AALGTFVAEVFVMAAFAFLPEKHEHGAATSYV---NLSSGMSNFLAP-----AAAVL- 377

Db 392 VGFAIGAVDSSMDPIMGVL--VDLRHVSYSVGSVAIADVAFCMGCVATIGPSAGGAAATAIG 449

QY 378 LPMNST-IGVY-IAYTALYLLAFLVCLGAFIRREGGFSAPTEKALNI 423

Db 450 FPMILMTIIGIDIAFAP-----LCEFLR-----SPPKKEKMAI 483

RESULT 12

PCT-US93-05704-4
Sequence 4, Application PC/TUS9305704
GENERAL INFORMATION:
APPLICANT: Edwards, Robert H
TITLE OF INVENTION: Vesicle Membrane Transport Proteins
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheldon & Max
STREET: 225 South Lake Avenue, Ninth Floor
CITY: Pasadena
STATE: California
COUNTRY: USA
ZIP: 91001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05704
FILING DATE: 19930611
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Farber, Michael B
REGISTRATION NUMBER: 32,612
REFERENCE/DOCKET NUMBER: 9067-1PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (818) 796-4000
TELEFAX: (818) 795-6321
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-05704-4

Query Match	5.6%	Score 123.5;	DB 5;	Length 515;
Best Local Similarity	21.8%;	Pred. No. 0.00042;		
Matches 89;	Conservative 63;	Mismatches 157;	Indels 99;	Gaps 23

```

QY 58 GLAALSAAMVS-----GVAAEITTPQKTLIGFVLMACPHFVLFVFGAGNAYGILL 110
    || || || || || || || || || || || || || || || || || || || || ||
Db 133 GLLRASKATVOLLNTPFLGLLTNNIGTPIPMFACF---CIMEISTVMFAR--SSSYAFLLI 188
    || || || || || || || || || || || || || || || || || || || || ||
QY 111 FYGIRGLAFLPLFLXSFYVIIHHNVRSSENSSALGWMYAVS-----VGIGVAGSYI- 161
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 189 ARSLQIGI-----SSCSVYAGMGLMSAVYTDDEBERGKPMGIALGLAMG 232
    || || || || || || || || || || || || || || || || || || || ||
QY 162 ----PSF--TIPMGEMGTLMLLACCFAGGYATMISLRHVKTIPGIMHNLTPREKFAEL 214
    || || || || || || || || || || || || || || || || || || || ||
Db 233 VLVGPPGVSYLEFGVKAPFLVLAALVLLDGAQLTLPLOPSRQOPESOKTP----- 285
    || || || || || || || || || || || || || || || || || || || ||
QY 215 SRAVTLTYTNINIFLSTIVRIINTLSLGFVIMPMKVEDELGFTTSEMLOVMAAFPTT 274
    || || || || || || || || || || || || || || || || || || || ||
Db 286 --LTTLLKDPILLAAAGSICFAN---MGIMAEPLPALPIMMEETMCSRKKOLGVAFILPAS 339
    || || || || || || || || || || || || || || || || || || || ||
QY 275 IF-----SNIFMGIVAEEKGMNRVIRWF--GCLGMAASSIAFYMDPOYGHNYMMAMIP-- 326
    || || || || || || || || || || || || || || || || || || || ||
Db 340 ISYLGITNIF--GILAHKMG-----RMICALLGMVIVGISLICIP--FAKNITGLIARNF 391
    || || || || || || || || || || || || || || || || || || || ||
QY 327 -AIALGTFVAFVMAAVALPELEKKHKGALISYV---NLSAGMSNLAP-----AIAVVL- 377
    || || || || || || || || || || || || || || || || || || || ||

```

Db 392 VGFLIGAWDSSMPIMGVL---VDLRHSVYGSYVAINDVAFCMGALGSPAGCAIKAKIG 449

RESULT 13

```

US-09-030-267-5
Sequence 5, Application US/09030267
Patent No. 6162632
GENERAL INFORMATION:
APPLICANT: Maloney, Peter C.
APPLICANT: Keletsu, Abe
APPLICANT: Zhong-Shi, Ruan
TITLE OF INVENTION: OXIT SEQUENCE AND ITS USE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
STREET: 3605 Glenwood Ave.
CITY: Raleigh
STATE: NC
COUNTRY: USA
ZIP: 27622
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,267
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,345
FILING DATE: 27-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: 5769-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2202
TELEFAX: 919 881 3175
INFORMATION FOR SEO ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-030-267-5

```

Query Match	5.68;	Score 123;	DB 4;	Length 418;
Best Local Similarity	20.68;	Pred. No. 0.00035;		
Matches 98;	Conservative 1 67;	Mismatches 139;	Indels 172;	Gaps 23

```

OY 17 MGVAIAIVFM-----TGQGFELAFASHYIK-SLGTPEASFAFTVLGAALSAWGV 71
Db 17 WEYLVAVLNCMTSGVSYMTLANKYKNTGSLAAVQTAFTLSGVIDAGSPGGYF 76
OY 72 AEITTPKMTLIGVL-----WCVFHV-----LFLVGLQANTGILLFTGI----- 114
Db 77 VDKGPRPLPMEFGAMVLAGTENGWDVSPALAYALTLAGAGYIV--YGIAMTANR 133
OY 115 -----RGIA-----YPLFYSEIYIVIHNVRENSSSALGWYWAYVSGIVAGSYIP 162
Db 134 WEPDKRGLAGSFGTAAGVGLGVLPELPLISSLVKE-----GVGAAFMYTG 178
OY 163 SFTIPINGEMKOTLALAFPCAGGVYAMISLRHKTPRGHHNHLPRKFAELBRAVTLX 222
Db 179 LI-----MGLITLILAFV-----IRFGF-----QQAKKOIV 206
OY 223 TNRNIFLSSIVRIINTLSLGEFAVIMPMMEVDELGTTSSEMLOWMAAF----- 271

```


Db 207 TDKDNSEGLMR-----TFQFWLMTAFSEVNGGLLVA 241

OY 272 -----FTTFSNT-----FMGIARMKG---MRYIRMGCGLGMA 304

Db 242 NSVYGRSGLAGVLLTIGVSIONLFGCGRPFGVSDIKYTKMSVY----FGINA 296

OY 305 SSIAFYMPQYGHNYMMAMPAIALGTFVAAFPMAAFPALEPKHKAISVYN----360

Db 297 VVALFPTIALGDVAFIML-AIAFTWGSY----ALFPTNSDIFGTAVSARVYGF 351

OY 361 --LSAGMSNF---LAPAIIVLLPFSTIGVYIATYATALLVLCAPT--RVEQP 409

Db 352 WAAKATASIFGGGIGAIAIATNF--GWNATAFLLTAITS--FIAPALATFVILPRGRP 403

RESULT 14

US-09-134-001C-5370

; Sequence 5370, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5370

; LENGTH: 499

; TYPE: PRF

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5370

Query Match 5.5%; Score 122; DB 4; Length 499;

Best local Similarity 21.3%; Pred. No. 0.00057;

Matches 92; Conservative 60; Mismatches 169; Indels 110; Gaps 19;

OY 17 WGIATAIVEMTGDFELAFISHYIKSLGFTTPABASFAFTLLYGLAALSAWVGVAAETIT 76

Db 36 WG---AFILPGD-----WIKQSG--PIASSIGIVIGALLMILIAVSGALVERKP 80

OY 77 POKTMLIGFVLMCVFHVLFIVEGLGOANYGLILFYG-----IRGLATVLPFLXSPVVI 130

Db 81 -----VSGAFAFESFLSGFRVYSFSSWFLTFGYVCVVALNATAFSLIKFLLPNV 131

OY 131 IHNVRENSSSALGWYAWVAVSGIVAGSYIPSF-TIPINGEMGTLMALAFCEPAGGVIA 189

Db 132 LNWGKLYTAA-----WDVYITELIATVLLIYFMILITIGASVSGSLQYFFCYA--MVL 184

OY 190 MISLRHVKTTPGHHM-NLTPREKPAELSRATVLLYTNRNIFLSSIVRIINTLSLFGFVIM 248

Db 185 VVALMFIQSFSSHFLSHLEPLASVDK-----WFGSLIMVS-----IA 225

OY 249 PMMFV--DELGFTTSEKLOVMAAFFTTIFSNIFWGIYAEMGMRYIRNFGCLGMASS 306

Db 226 PMAYVGFNDINIPOTAEER-----NESPKNKTEKLIVSILASLTYYVVALTYTGMISTQATS 280

OY 307 LAFYVMPQYGHNYMM-----AMIPALGTFA--AFVP-----339

Db 281 L-----NGNIMLTGAVTODAFEGTIGLAVLAVAILIMIGITGLNGFLMSSSRLLFSMG 331

OY 340 -----MAVFPALPEKHKGAISVYNLSAGMSNFLAPAIIVLLPW--FSTIGVIAVTA 392

Db 332 RSGIMPTVFESKLSKHTPTVAIIIFVA--VSLIAPWLGRTALTWIYDMSTGVSIAVFI 389

OY 393 LYLLAFVLCAP 403

Db 390 TCI SATKLF 400

RESULT 15

US-09-134-001C-3949

; Sequence 3949, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3949

; LENGTH: 412

; TYPE: PRF

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3949

Query Match 5.5%; Score 121.5; DB 4; Length 412;

Best local Similarity 19.0%; Pred. No. 0.00049;

Matches 74; Conservative 81; Mismatches 168; Indels 67; Gaps 17;

OY 39 YKIS-LGFTP-----AESFAFTLYGLAALSAWVGVAAETITPOKTMILGFVLMCV 90

Db 47 YILSYNLEPTVYSIAIATSFASFISHPADATNFVIGLKKFGSKLVYISGILL--A 104

OY 91 FHVLFVFGIGGOANYGLILLEYGIR-GLAY-PLFLYSFVYIHNVRSENSSSALGWYA 148

Db 105 FLSFLVIMFPASP--IIFSAIMGLIAVSPIM-----VIMLSVDERNRGRKMGVYVE 157

OY 149 VYVGVAGVASYIPSTIPIMGEMGTLMALAFCEPAGVYAMISLRHVKTTPGHHMLTPR 208

Db 158 SWLGL-LVGMVIMNLIKHPTRFAFLMALVVLIAWLVLYFVNI-----MLTYNTRK 209

OY 209 EKRAELSRATVLLYTNRIPL-----SSIVRIINTLSLFGFAYIMPMFVDELGFTT 260

Db 210 PVKQOLKQIVD--TORHLILPFGILLGGAIALVALPILKVIATOVYK---VSTVGYTV 263

OY 261 SEMIQ---VMAAFETTIFSNIFWGIYAEMGMRYIRNFGCLGMASSIAFYMPQYF 316

Db 264 AIIIGIGCAFSLFLSKIIDN-----NSKGFMYGVIESGFI-----LYTILIF 307

OY 317 G---HNYYMMAMIPALGTFAVFPMAAVFPA--LEPKHKAISVYNLSAGMSNFLA 370

Db 308 GLSTITNITYIYMAIGLFIGLMYGLIPAWNTFMAGHINPNEDEETWGVENSVOGFGSMIG 367

OY 371 PAIAVLLPWFSTIGVYIATYATALLVLCAPT--RVEQP 409

Db 368 PLVGGILTPTNNIMNTFESAMIFLALAV 397

Search completed: March 13, 2003, 16:42:19

Job time: 13.5828 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 11.9338 Seconds
(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNKQWLGILPLHLIMGYI.....RVGEQGFSSAPYTEKALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2211	100.0	424	US-09-802-208b-5	Sequence 5, Appl1
2	580.5	26.3	445	US-09-738-626-3627	Sequence 3627, Ap
3	143.5	6.5	595	US-09-738-626-6943	Sequence 6943, Ap
4	142	6.4	463	US-10-122-466a-11	Sequence 11, Appl
5	139	6.3	417	US-09-815-242-10592	Sequence 10592, A
6	138.5	6.2	449	US-09-738-626-6555	Sequence 6555, Ap
7	136.5	6.0	425	US-09-815-242-10467	Sequence 10467, A
8	132.5	5.9	378	US-09-738-626-6793	Sequence 6793, Ap
9	130	5.9	701	US-09-738-626-6196	Sequence 6196, Ap
10	129.5	5.9	453	US-09-738-626-6877	Sequence 6877, Ap
11	128	5.8	450	US-09-738-626-4815	Sequence 4815, Ap
12	126	5.7	429	US-09-922-501-10	Sequence 10, Appl
13	125.5	5.7	407	US-09-815-242-11180	Sequence 11180, A
14	124.5	5.6	448	US-09-738-626-6795	Sequence 6795, Ap
15	123.5	5.6	472	US-10-024-623-29	Sequence 29, Appl
16	123	5.6	663	US-09-815-242-14080	Sequence 14080, A
17	122	5.5	663	US-09-815-242-10070	Sequence 10070, A
18	121.5	5.5	462	US-09-738-626-5120	Sequence 5120, Ap
19	121.5	5.5	467	US-09-738-626-3974	Sequence 3974, Ap

20	121	5.5	450	10	US-09-741-669-385	Sequence 385, App
21	120.5	5.5	427	10	US-09-922-501-2	Sequence 2, Appl1
22	120	5.4	426	10	US-09-922-501-4	Sequence 4, Appl1
23	120	5.4	551	9	US-09-738-626-4431	Sequence 4431, Ap
24	119.5	5.4	454	10	US-09-729-674-140	Sequence 140, App
25	119.5	5.4	469	10	US-09-925-302-627	Sequence 627, App
26	119.5	5.4	534	9	US-09-981-947A-4	Sequence 4, Appl1
27	119.5	5.4	649	10	US-09-815-242-5333	Sequence 5333, Ap
28	119.5	5.4	662	10	US-09-815-242-12343	Sequence 12343, A
29	119	5.4	526	9	US-09-738-626-5712	Sequence 5712, Ap
30	118.5	5.4	451	10	US-09-815-242-5079	Sequence 5079, Ap
31	118	5.3	513	9	US-09-738-626-6860	Sequence 6860, Ap
32	116	5.2	476	9	US-09-946-673-2	Sequence 5741, Ap
33	115.5	5.2	658	10	US-09-815-242-11824	Sequence 11824, A
34	114.5	5.2	395	10	US-09-815-242-13892	Sequence 13892, A
35	114	5.2	497	9	US-09-738-626-6734	Sequence 6734, Ap
36	113.5	5.1	464	9	US-10-024-623-28	Sequence 28, Appl
37	113.5	5.1	549	9	US-09-946-673-2	Sequence 3926, Ap
38	113.5	5.1	549	10	US-09-946-673-2	Sequence 2, Appl1
39	113	5.1	372	9	US-09-860-670-106	Sequence 106, App
40	112.5	5.1	391	10	US-09-815-242-11912	Sequence 11912, A
41	112.5	5.1	468	10	US-09-815-242-5144	Sequence 5144, Ap
42	112.5	5.1	487	9	US-09-738-626-4875	Sequence 4875, Ap
43	112.5	5.1	501	9	US-10-051-909-38	Sequence 38, Appl
44	112	5.1	403	9	US-09-738-626-6610	Sequence 6610, Ap
45	112	5.1	417	10	US-09-815-242-10165	Sequence 10165, A

ALIGNMENTS

RESULT 1

US-09-802-208b-5

Sequence 5, Application US/09802208B

Publication No. US20030041352A1

GENERAL INFORMATION:

APPLICANT: Parrott, Wayne

APPLICANT: Lafayette, Peter

APPLICANT: Kane, Patrick

TITLE OF INVENTION: Arabidol or Ribitol As Positive Selectable Markers

FILE REFERENCE: UGA-853R

CURRENT APPLICATION NUMBER: US/09/802, 208B

CURRENT FILING DATE: 2001-03-08

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 424

TYPE: PRT

ORGANISM: Escherichia coli

US-09-802-208b-5

Query Match 100.0%; Score 2211; DB 9; Length 424;

Best Local Similarity 100.0%; Pred. No. 2.5e-181;

Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSRRNKQWLGILPLHLIMGYIAIVFMTGDFELAFISHYKSGFPAPASPAFTLVGLA 60

1 MSRRNKQWLGILPLHLIMGYIAIVFMTGDFELAFISHYKSGFTAPASPAFTLVGLA 60

61 AALSAMVSGVAAEITTPQKTMILGFEVLMCVHFVLFVFGGLGQANYGLLFGYGRGLAYP 120

61 AALSAMVSGVAAEITTPQKTMILGFEVLMCVHFVLFVFGGLGQANYGLLFGYGRGLAYP 120

61 AALSAMVSGVAAEITTPQKTMILGFEVLMCVHFVLFVFGGLGQANYGLLFGYGRGLAYP 120

121 LFLYSLFVLIHNVRSNSSLGAGMYAVYVSGVAGVSTPSTTIPKEMGLMLALA 180

121 LFLYSLFVLIHNVRSNSSLGAGMYAVYVSGVAGVSTPSTTIPKEMGLMLALA 180

121 LFLYSLFVLIHNVRSNSSLGAGMYAVYVSGVAGVSTPSTTIPKEMGLMLALA 180

181 FCPAGVYIAIISLRHVTPGPHMLUPREKFAELSRVATLLYTRNFSLSIYRIINTLS 240

181 FCPAGVYIAIISLRHVTPGPHMLUPREKFAELSRVATLLYTRNFSLSIYRIINTLS 240

181 FCPAGVYIAIISLRHVTPGPHMLUPREKFAELSRVATLLYTRNFSLSIYRIINTLS 240

241 LFGFVAVIMPVAVDELGFTTSEMLOVAAFEFTTIFSNIFGVAIEKMGMRVIRMGCL 300

Db	241	LEGRAVIMPMFVDELGPTTSEMLQVMAAFPEFTTIFSNIFMGIVAEKMGMRIVIRMFGL	3000
Qy	301	GMAASSLAFYMPQYEGHNVMAMIPALGLTEVFAAFVPMAAVFPALPEPHKGAISVYN	3606
Db	301	GMAASSLAFYMPQYEGHNVMAMIPALGLTEVFAAFVPMAAVFPALPEPHKGAISVYN	3606
Qy	361	LSAGMSNFLAIVVLLPMFSTIGVVIATATLLFLFVLCAPFRVQPFSSAPVTEKA	4200
Db	361	LSAGMSNFLAIVVLLPMFSTIGVVIATATLLFLFVLCAPFRVQPFSSAPVTEKA	4200
Qy	421	LNIS 424	4200
Db	421	LNIS 424	4200

```

RESULT 2
US-09-738-626-3627
: Sequence 3627, Application US/09738626
: Publication No. US20020197605A1
GENERAL INFORMATION:
: APPLICANT: NAKAGAMA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKOI, HARGHIO
: APPLICANT: TATEISHI, NAOCO
: APPLICANT: SENOH, AKIHRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 3627
: LENGTH: 445
: TYPE: prt
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3627

```

Query Match	26.3%	Score 580.5	DB 9	Length 445
Best Local Similarity	30.4%	Pred. No. 4.8e-42		
Matches 131	Conservative 89	Mismatches 188	Indels 23	Gaps 8

[illegible][illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	6.5%	Score 143.5	DB 9	Length 595
Best Local Similarity	20.5%	Pred. No. 0.00014		
Matches 105, Conservative	70	Mismatches 154	Indels 183	Gaps 26

QY	20	IAIVAFPMGDOEFELAFISHYTKSLGTFPAEASFPEFTLYGLAALASAVSGVAEIIIPQK	79
Db	69	IVLAVVWGIGDFKDSFTMFASSALSAVVDNLGMFFILFG-----TVFVFPIVIAASKFG	123
QY	80	TMIG-----FVLMCVFHVLFVEGLGQANVGLILFFYGRGLAVPLETFSPIV-	129
Db	124	TIRGRIDEAPEPTVSM-----ISMFPAGM-----GIGLMFY-----CTTEPLETFRRNVPG	172
QY	130	---IHHVRSSESSALGW---MAVYSVIGVAGSY-----IPSFPIPIGKNGTL	175
Db	173	HDEHNVGAASTMTFHHHTLHPMALIYAL-VGLAIAYSTFRPRGRQLSSARVPLIGEGAE	231
QY	176	-WL-----ALAP-----CEAG-----GVIAITSLRHV	196
Db	232	GWLKLIDILAIITVETGACSLGALIGAGLSAANIETEDPSDMTIVGIVSVLTAFI	291
QY	197	KTPGHHNLLPREFAELS---RAVTLILYNNRNLFLSSIVRII-----	236
Db	292	-----FSAISGVGKIQYL-SNANVYLAALALAFVYVYGPVVISLNLPGS	336
QY	237	---NTLSLGFPAVIMPMFVVELGFTTSBEMIQVMAAFETTIFIS-NIPMGIVAEKMGMRV	293
Db	337	IGNLTSNFPQAGRTAASADG---TAGEWGSLWTFIVWAMWISPSVPVGMFLARISGRS	393

Db 309 SVFAIAFVTRNTLLISVLKGLTKLALDPIMVTVTKHAPKALGTTLSAIVFGMSG 368
QY 367 NEFLAPAIIVLLPWFSTIGVIAVYATYALYLAFLVLCAP 403
Db 369 SILAPVVTGYLADPAGSMQVCF-YLSCVLLVYGLLAF 404

RESULT 6

US-09-738-626-6555
; Sequence 6555, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKOTO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6555
; LENGTH: 449
; TYPE: PRF
; ORGANISM: *Corynebacterium glutamicum*
US-09-738-626-6555

Query Match 6.3%; Score 138.5; DB 9; Length 449;
Best Local Similarity 21.0%; Pred. No. 0.00027;
Matches 88; Conservative 53; Mismatches 125; Indels 153; Gaps 21;

QY 7 QWLG--LPLHLIMGY---IAIAVPMGDFELAFLSHYIKSLGFTPAE-ASFA----- 53
Db 127 EMFGMMLPVGFVGGFATAAVGTALBESSGAEAM-----SLGFTSATVGTFAAIVGCI 180
QY 54 FTLYGL-----AAALSANV-----SGVVAELITTPQKTMILGIVLMCVHVLFLVFGLGQAN 104
Db 181 FTWGIKKGTAAAPQGLPMDLNSGYIDKLSDRPS-----IGKASTN 222
QY 105 YGLLLFVGIRGLAYPLFLYSFTVVIITHNVRSESSALGMWMAVYSGVIGVAGSYIPSE 164
Db 223 PSAL-----EPLALHTGIILLTVN-----AYSINOML-GSMPTIV 257
QY 165 TPIIMGEMGLMLALAFCEPAGVIAIMISLRHVKTGPHMNLTPREKFAEL-SRAVTTLLYTN 224
Db 258 QIDP-----FAMSFVAGIYGMGIMRLKPREYL-----DRDTVNSVSGAAT----- 298
QY 225 RNIFLSSIVRIITLSLFGFAVIMPMFVDELGFTTSEMLOWMAAFPTTIFSIIFGCIY 284
Db 299 -----DYLIAFGIASIAPAIAD-----YMWPLVVLFLVGLTYCCFFP--- 336
QY 285 AERKMWRVIRWFGCLGMAASSLAIFYMPOYFGHNWMAIPALIGTFVAAPVPMAYF 344
Db 337 -----FWAPRPFGEKMLERAI--FGKGMATAAATAVATGIAL 370
QY 345 PALEPKHKAISVYNLS-----AGMSNFLAPAIAYVL-----LPWFSTIGVIAV 390
Db 371 KIYDPKIKSGALNBYGVAIYGAPFEIGMT-IAP-IAYLAGFTMGMSLVAIAYIF 427

RESULT 7

US-09-815-242-10467
; Sequence 10467, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10467
; LENGTH: 425
; TYPE: PRF
; ORGANISM: *Escherichia coli*
US-09-815-242-10467

Query Match 6.2%; Score 136.5; DB 10; Length 425;
Best Local Similarity 20.3%; Pred. No. 0.00038;
Matches 93; Conservative 63; Mismatches 136; Indels 161; Gaps 23;

QY 8 WLGLPLHLIMGYIAIAVPMGDFE---LAFLSHYIKS-LGFTPAEASFATYIGLAAL 63
Db 41 WLGY-----VE---DGEDFMIRFYLIHIIKADLGITDQATLIGVAFIARPI 85
QY 64 SAWVSGVVAELITTPQKTMILGIVLMCVHVLFLVFGLGQANTGLILFPYIGIRGLA--YP 120
Db 86 GGGEFGAMADKY-GRKPM-----HMAIF-----IYVGT-----GLSGIATNLVM 125
QY 121 LFLYSFTVVI-----IHNVRSESSALGMWMAVYSGVIGVAGSYIP----- 162
Db 126 LAYCRFLVIGLMSGEVACASTAYAVESPKNLOKASAFVYSGSVGIIIAQIIPORAEV 165
QY 163 -----SFTPIIMGEMGLMLALAFCEPAGVIAIMISLRHVKTGPHMNLTPREKFAEL 217
Db 186 YGNRNSPFIGLPLVILVLM-----RKSAPESQEW 215
QY 218 VTLTYTRNIFLS-----SLVRIITNLSLFG-----FAVIM 249
Db 216 IEDKYKDKSTFLSVFRKPHLSISNIVELVCECFGANMPINGLLPSYADNGVNTVVIS 275
QY 250 IMFVDELGFTTSEMLOWMAAFPTTIFSNIFMGIVAEKGMWIRWFGCLGMAASSLAF 309
Db 276 LMTIAGL-----TLNGTIFFGVGDIGVKKAF-----VGLITSTI-- 313
QY 310 YNMPDYFGHNWMAIPALIGTF-----VAAPVPMAYPALEPKHKAIS-VYNLSA 363
Db 314 FLCPLEFTISVKNSSLLIGLCPGLMFTNLGIAGLVP-KFIYDTPPTKIRGLGTGLIYNLGA 372
QY 364 --GMSNFLAPAIIVLLPWF--STIGVIAVYATYLL 396

Query Match	5.9%	Score 129.5	DB 9	Length 453
Best Local Similarity	21.3%	Pred. No. 0	0016	
Matches 97; Conservative	75;	Mismatches 180;	Indels 103;	Gaps 22

RESULT 11
US-09-738-626-4815
; Sequence 4815, Application US/09738626
; Publication No. US20020197605A1
; Patent Information

1 APPLICANT: NAKAGAMA, SAOTOSHI
2 NAKAGAMA, SAOTOSHI
3 MIZOGUCHI, HIROSHI
4 APPLICANT:
5 APPLICANT:
6 APPLICANT:
7 APPLICANT: HAYASHI, MIKIRO
8 HAYASHI, MIKIRO
9 APPLICANT: OCHAI, KEIKO
10 OCHAI, KEIKO
11 APPLICANT: YOKOI, HARUHIKO
12 YOKOI, HARUHIKO
13 APPLICANT: TATEISHI, MAKOTO
14 TATEISHI, MAKOTO
15 APPLICANT: SENOH, AKIHIRO
16 SENOH, AKIHIRO

Query Match	5.8%	Score 128	DB 9	Length 450
Best Local Similarity	21.2%	Pred. No. 0.0022		
Matches 90	Conservative 62	Mismatches 163	Indels 110	Gaps 21

RESULT 12
US-09-922-501-10
; Sequence 10, Application US/09922501

```

1  TITLE OF INVENTION:  2, 5-DKG./PERMETHAS
2
3  FILE REFERENCE:  P-SR 4877
4
5  CURRENT APPLICATION NUMBER:  US/09/922,501
6
7  PRIOR FILING DATE:  2001-08-03
8
9  PRIOR APPLICATION NUMBER:  US 09/633,294
10
11 PRIOR FILING DATE:  2000-08-04
12
13 PRIOR APPLICATION NUMBER:  US 09/677,032

```


; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07

```

? PRIOR APPLICATION NUMBER: JP 00/280988
?
? PRIOR FILING DATE: 2000-08-03
?
? NUMBER OF SEQ ID NOS: 7059
?
? SOFTWARE: Patentln ver. 3.0
?
? SEQ ID NO 6795
?
? LENGTH: 448
?
? TYPE: prt
?
? ORGAINSH: Corynebacterium glutamicum
?
? JS-09-73b-626-6795

```

Query Match	5.6%;	Score 124.5;	DB 9;	Length 448;
Best Local Similarity	22.2%;	Pred. No. 0.0043;		
Matches 102; Conservative	68;	Mismatches 163;	Indels 127;	Gaps 23;

```

OY 29 DGFELAFASHIKL-----GTPPEASFAFTVLGLAALASWGVAAETITPOKMTLG 84
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 41 DAMVGILSFPMALATHWGLSPETSLDLSIGVGMATGASLGGLLADNIGRQ----- 95
OY 85 FVLMCVHVLFVLVGLG-----QANYGLLFLFYINGLAY-----PLEL 123
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 96 -----VFALSLFVGVATGASALSVSALMLAFVYVGLDGLAELPAVSTLISESPRKY 150
OY 124 YSFIVVLIHNVRSESSALGGMVAVYVSGVGAVGSIIPFTPIPKGEMKTYLLALAFCE 183
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 151 RGRWVLI-----EATPMALGWMALAVGTFFVAGS-----DNQWRN-ALALGC 192
OY 184 AGGVIAIMI-----SLRHVYTPGCHMIN-----LTREKFEALSRAY----- 218
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 193 VPATVAVYVRLGDPESVRLEKKR-HDEAVALVVSFEAAANAEKKAADTTAVYHDNA 251
OY 219 -----TLTYNNRIFLSSIVRIILNLSLGFRAVYIPMKMFVDELGTTSEWLDQVAAF 271
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 252 EGSVSIWSAALRKKTVALMTWIEFIN-LSYGAETWIPSLIVAD-GFTLVK-----SFO 303
OY 272 FTTI-----FSNIFMGIVAEKMGMMRYIRMGCGGMASSLAFYVMQYFGRHYWMA 323
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 304 FTLLIITLAOLPGYVAAH--LIEWGRKRSILATP--LVGSAISALXYGL-----ANEMQI 355
OY 324 MIPALATGF-VAAFVPMAAVFPALBEKH-----KGAASIVYNLSAGMSNFLAPALAVYL 377
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 356 LVACLLSFFMIGAMGALYALGAPELAYPTNNGTGTGA-----AGGR-IASITAPLI 407
OY 378 LPWFSTIGVATYATYALYLAF-----VLCAFLRVDPGSSSA 414
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 408 VPPVIAEGPITALFATATAIAIAAFTLLPDEKKGSLA 447

```

RESULT 15
US-10-024-623-29

```

: Sequence 29 Application US/10024623
: Publication No. US20020167524A1
:
: GENERAL INFORMATION:
: APPLICANT: Curtiss, Rory A.J.
: TITLE OF INVENTION: 8099, 46455, 54414, 53763, 67076, 67102, 44181,
: TITLE OF INVENTION: 67084FI, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF
: TITLE OF INVENTION: USE THEREOF
: FILE REFERENCE: MNI-21ACP
: CURRENT APPLICATION NUMBER: US/10/024,623
: CURRENT FILING DATE: 2001-12-17
: PRIOR APPLICATION NUMBER: US 60/256,240
: PRIOR FILING DATE: 2000-12-15
: PRIOR APPLICATION NUMBER: US 60/256,588
: PRIOR FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: US 60/258,028
: PRIOR FILING DATE: 2000-12-21
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 29
: LENGTH: 472
: TYPE: PR1
: ORGANISM: Escherichia coli
: US-10-024-623-29

```

Query Match	5.68;	Score 123.5;	DB 9;	Length 472;
Best Local Similarity	20.88;	Pred. No. 0.0055;		
Matches 98;	Conservative 73;	Mismatches 163;	Indels 137;	Gaps 27;

```

OY 19 YIAINPMTGGGFEELAFSLSHIKSLGTPAPASAEFTLYGAAALSMVAGVAEITTPQ 78
Db 24 FVSVAANAAGLLFGL-----DVGYAGALPPTTHFVLTSLQEMV-----V 65
OY 79 KTMILGFVLMCVFHVLELVFGGQANT---GLILPEYGRGLAYPLEVYSFVLIITHNV 134
Db 66 SSMMLGAIGALFNG-WLSFRLGR-KYSLMAGALLFVLGSIGSAFATSVEMLEIAARV--- 120
OY 135 RSENNSSALGMVYMYVSVGIGVAGSYIPSEFIP-IGEMBG-----TLML 177
Db 121 -----VLGIAVGA-----SYAPPLYSEMASENVRGKISMISYOLMLVTLGI 164
OY 178 AL-----AFCFAG-----GVIAISL-----RHYKTPRGHML 205
Db 162 VLAFSLSDFAFYSQGWMMRAMGLVLAIPVLLILVPLPNSPRMLAEKGRHIEAEVRLML 221
OY 206 -----TPREKFAELSRV-----TLLYNRN-----IFSSIVRIINTLSLGFAYIM- 248
Db 222 RDTSEKAREEENETRESLKLQSGMALEFKIRNNYRAVLEMLLQAOQOFT--GMNIMY 279
OY 249 ---PMWEVDELGETTSE-----WLOVMAAEFFTTTFESNFMGIVAEMKMMVRIRPFGLG 301
Db 280 YAPRLF-KMAGFTTTEOOMLITLVYVGLTEMFAT-FIAVF---IYDKKGRPAK-IGFSV 333
OY 302 MAASSLAFYVMPDYFGHN-----YMMAM-IPAIALGTEVAAPVMAAVF-PALER-KHK 352
Db 334 MALGLTVLYGCLMOPDNCTAGSSGLSWLSVGTMTMCIAGYASAPVYMILCSEIOLPKR 393
OY 333 GAISIVYVLSAGMSNLEPALAVVLLPREFSTIGVIVATYALXLLAFVLCAF 403
Db 394 DFGITCSTTMMVSMNIIIGAFITFLLLDSIGANGAFWMLTALN-LAEVGITF 443

```

Search completed: March 13, 2003, 16:51:26
Job time : 14.9338 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 14.7417 Seconds
(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNQWMLGPLHLINGYL.....RVQPGSSAPVTEKALNIS 424

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	1917	86.7	427	2	S78599
2	821	37.1	435	2	H69607
3	709.5	32.1	414	2	D68895
4	181	8.2	456	2	D83117
5	180.5	8.2	456	2	A71501
6	176	8.0	456	2	A10384
7	174	7.9	388	2	A37838
8	173.5	7.8	444	2	G70042
9	168.5	7.6	457	2	E71614
10	164.5	7.4	386	2	G64494
11	162.5	7.3	426	2	AF6F34
12	161.5	7.3	400	2	C70082
13	159.5	7.2	428	2	AD0675
14	158.5	7.2	387	2	C86640
15	158.5	7.2	455	2	F86573
16	158.5	7.2	455	2	A72050
17	158.5	7.2	456	2	G81660
18	155.5	7.0	419	2	A83133
19	155	7.0	412	2	E84208
20	154.5	7.0	380	2	S74023
21	153	6.9	405	2	T44249
22	153	6.9	405	2	AE0990
23	152	6.9	396	2	F69813
24	151.5	6.9	428	2	E70078
25	151	6.8	413	2	B84085
26	150.5	6.8	409	2	D75358
27	150	6.8	401	2	AH1584
28	150	6.8	437	2	H83175
29	149.5	6.8	425	2	A90055

30	149.5	6.8	567	2	C75340	probable L-lactate
31	148.5	6.7	402	2	A83398	probable MFS trans
32	147	6.6	401	2	A81231	antibiotic resista
33	147	6.6	431	2	D81282	antibiotic resista
34	146.5	6.6	436	2	G72299	probable efflux pr
35	145.5	6.6	381	2	F71078	conserved hypothet
36	144.5	6.5	378	2	A99263	hypothetical prote
37	144	6.5	429	2	A83574	transpore membrane
38	143.5	6.5	420	2	AG0958	multidrug resistan
39	143	6.5	456	1	C64772	probable membrane
40	143	6.5	743	2	AE3145	probable transport
41	143	6.5	747	2	F98142	hypothetical prote
42	142.5	6.4	413	2	B87451	iron(III) ABC tran
43	142.5	6.4	454	2	AD0556	membrane protein,
44	142	6.4	406	2	AH0563	hypothetical major
45	142	6.4	463	1	A35620	fosmidomycin resis
						peptide transport

ALIGNMENTS

RESULT 1	
S78599	probable D-ribulose transporter - Klebsiella pneumoniae
N:Alternate names: ribitol transporter	
C:Species: Klebsiella pneumoniae	
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 08-Oct-1999	
C:Accession: S78599	
J:Henel, H.; Turgut, S.; Schmid, K.; Langelier, J.W.	
R: Bacteriol. 179, 6014-6019, 1997	
A:Title: Substrate recognition domains as revealed by active hybrids between the D-ar	
A:Reference number: S78599; MUID:97464425; PMID:9324246	
A:Accession: S78599	
A:Molecule type: DNA	
A:Residues: 1-427 <HEU>	
A:Cross-references: EMBL:AF045244; EMBL:U07127; NID:92905642; PIDN:AAC26496.1; PID:92	
A:Experimental source: strain KAY2026	
C:Genetics:	
A:Gene: rbtR	
Query Match	86.7%; Score 1917; DB 2; Length 427;
Best Local Similarity	85.6%; Pred. No. 1.2e-134;
Matches 363; Conservative 28; Mismatches 29; Indels 4; Gaps 1;	
QY	1 MSRRNQWMLGPLHLINGYLAIVEMTGDFELAFSHYIKSGFTPARSPAFLLYGLA 60
DB	1 MSYNNQWYGLPLNLIMGWALNAVMTGDFELAFSHYIKALGFTPAQNSFAFTLYGLA 60
QY	61 AALSAVSGVVAEIIIPKTKMLIGFVLMCVFHVLFVFGIGQANGLILFFGIRGLA 120
DB	61 AALSAVSGVVAEIIIPKTKMLIGFVLMCVFHVLFVFGIGQANGLILFFGIRGLA 120
QY	121 LFLYSFIVYIHNVRSESSSALGWYAVYSGVIGAGSYIPSTPIPGEMKTLALA 180
DB	121 LFLYSFIVYIHNVRSDSSSALGWYAVYSGVIGAGSYIPSTPIPGEMKTLALA 180
QY	181 FCGAGGVIMISLRHVKTGCHMNLTPRKKFALSAVLLTYNNRMLISSTVRIINTLS 240
DB	181 FCGAGGVIMISLRHVKTGCHMNLTPRKKFALSAVLLTYNNRMLISSTVRIINTLS 240
QY	241 LFGFAVIMPMMEVDELGFTTSEMLQVMAAFETTFISNIFMGVIAKMGMRVIRMGGL 300
DB	241 LFGFAVIMPMMEVDELGFTTSEMLQVMAAFETTFISNIFMGVIAKMGMRVIRMGGL 300
QY	301 GMAAASLAFYVMPQYRGHNYWMAIDALGTFVAAFPVMAAVPALDEPKHKAISVYN 360
DB	301 GMAAASLAFYVMPQYRGHNYWMAIDALGTFVAAFPVMAAVPALDEPKHKAISVYN 360
QY	361 LSAQGNFLAPATAVLLPWFSTIGVIAVTAIYLAFVLCARIRVQGGSS----AVY 416
DB	361 LSAQGNFLAPATAVLLPWFSTIGVIAVTAIYLAFVLCARIRVQGGSS----AVY 416
QY	417 TEKA 420

Db 421 TANA 424

RESULT 2
H69607

alpha-ketoglutarate permease csbx - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: H69607

R:Kunst, S.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth

C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.: Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MID:98044033; PMID:9384377

A:Accession: H69607

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1435 <KUN>

A:Cross-references: GB:299118; GB:AL009126; NID:g2635200; PIDN:CAB14736.1; PID:g2635241

A:Experimental source: strain 168

C:Genetics:

A:Gene: csbx

C:Superfamily: hypothetical protein c0103

Query Match 37.1%; Score 821; DB 2; Length 435;

Best Local Similarity 36.5%; Pred. No. 1.6e-53;

Matches 146; Conservative 89; Mismatches 161; Indels 4; Gaps 2;

9 LGPLHLIMGYIAIAVMTDGFELAFLSHYIKSLGFTPAEASFAFLYGLAALASAWS 68

Db 14 IGIRPMHWGFLGVLFEMDGLGQGLSPFLVDHGLSMQSSLPFTMYIAVTISMWS 73

QY 69 GVAEITTPDKTMLIGVLCVFNVLVFLVGLQANGLLLEFGYGLAVPLFLYSFIV 128

Db 74 GFEVQTMGPKRTVGLAFILGSAFICGAIIPMYRPAALGSLALRGGLYPLFAVSFLV 133

QY 129 VIHNHNSSENSSALGMYMAVSVIGVAGSYIPSPFTIPIMGEMGTMLALACFAGVI 188

Db 134 WVSYSQNLGKAVGFWFMFTGCLNVLCGFYSYAVPAFGEINLTLMSALLFVAAGSL 193

QY 189 AMISLRHVKTGPHMNLTPREKFAELSRVATLLYTNRNIFLSSIVRIINLTLSEFVIM 248

Db 194 ALFFNNDKFPPIKOD---QPKMKELSKAFITMFPNPKVIGGVKTINMIGOFGRIFL 250

QY 249 PMMFVDELGFTISEMLQVNAAFPTTIFSNIFMGVIAEKGMKMRVIRWFGCLGMAASSLA 308

Db 251 P-TYLARYGYSVEWLMQIMGTLEFVNIVFNIIIGAVGDKLGMHNTVMWFGVCGGIFTLA 309

QY 309 FYVMPQYFGHNMYMMAMPALALGTFAVAPVMAVPPALPEPKHGAISVYNLSAGMSNF 368

Db 310 LYTTPOLIGQVWMLIACCYGAALAGVYPSLALPTLPAIDKGAAMSVNLNIGFGLCAF 369

QY 369 LAPAIVALLPWFSTIGVAVIATVLAFLVLCATIRVEQ 408

Db 370 IAPGIVSLFIRPGAGGVIMFALILFFSAFLTRFLITSE 409

RESULT 3

D69895

conserved hypothetical protein yaab - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-May-2002

C:Accession: D69895

R:Kunst, S.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.: Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MID:98044033; PMID:9384377

A:Accession: D69895

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1414 <KUN>

A:Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13747.1; PID:g26342

A:Experimental source: strain 168

C:Genetics:

A:Gene: yaab

C:Superfamily: hypothetical protein c0103

Query Match 32.1%; Score 709.5; DB 2; Length 414;

Best Local Similarity 33.6%; Pred. No. 2.8e-45;

Matches 139; Conservative 88; Mismatches 172; Indels 15; Gaps 4;

9 LGPLHLIMGYIAIAVMTDGFELAFLSHYIKSLGFTPAEASFAFLYGLAALASAWS 68

Db 5 IGIRPMHWGFLGVLFEMDGLGQGLSPFLVDHGLSMQSSLPFTMYIAVTISMWS 64

QY 69 GVAEITTPDKTMLIGVLCVFNVLVFLVGLQANGLLLEFGYGLAVPLFLYSFIV 128

Db 65 GVCLEARGAKRTMFMGLFVIGTAIVFGFBDLNPVWYVYVGLGYPLFAVSFLT 124

QY 129 VIHNHNSSENSSALGMYMAVSVIGVAGSYIPSPFTIPIMGEMGTMLALACFAGVI 188

Db 125 WVIKTPQSLSTRVAGFMAIYCLGMFPGAMVSSYIAKAFGLYNTLMSSIFWVGLCAF 184

QY 189 AMISLRHVKTGPHMNLTPREKFA-ELSRVATLLYTNRNIFLSSIVRIINLTLSEFVIM 247

Db 185 ALF----INKDREKKRKRKSEFAEELKGVTLITFTRPVLTGIIIRINSIGYGPVE 240

QY 248 PMMFVDELGFTISEMLQVNAAFPTTIFSNIFMGVIAEKGMKMRVIRWFGCLGMAASSL 307

Db 241 LPMHMAOH-GISTNVWLMQIMGTLEFVNIVFNIIIGAVGDKFGMKNTVMWFGVCGGIFTL 299

QY 308 AFYVMPQYFGHNMYMMAMPALALGTFAVAPVMAVPPALPEPKHGAISVYNLSAGMSN 367

Db 300 LLYTAPYFSGSLAVSVIGFINGGLLAGYVPGIATVPTVYAGDKGAMSVNLNIGFGLSA 359

QY 368 FLAPAIIVALLPWFSTIGVAVIATVLAFLVLCATIRVEQPGFSSAPVTEKAL 421

Db 360 FVSPALAMLEIGLVAGGVWIMFALVILASAVLTKIHIDE-----EKAV 404

RESULT 4

D83117

probable MFS transporter PA4233 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: D83117

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.D.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MID:20437337; PMID:10984043

Matches 105; Conservative 71; Mismatches 171; Indels 123; Gaps 58

Db 165 GLOALFWCIAITALLGLIVITLTPVPSANSHVLNRESSMVKGSVKLHNSRLKLNGCIM 224

```
Db 225 CILHLLMSSE-VALPOMMAN-AGLAPA---OHMVYVLTMLVSPFAAVVPTIYAEMKRRM 279
OY 278 -NIMGVIAEKMGMVRVIRMFEGCLMAASSLAIFYMPOYGHNMWMAIRPALACTPAA 336
Db 280 KOVPMGCA-VLFAEVVLMFA-----GODLMT-ITACVQL--FEIA 317
OY 337 FVPAVAEPALPEK-----HGGAISVYNLSAGNSNFIAPAIVALLPW-FSTIGVITAY 390
Db 318 FNVWEALPILSLISKESPAKGTAMGITS-----TSQFIGVAIGSLGIMFGLGADNVF 373
OY 391 TALYLAFVLCAF-IRVEQPGFSSA 414
Db 374 AAGAIIALMPFAVSVTQMDEPPYVSS 398

RESULT 7
A37838
nora protein - Staphylococcus aureus
C:Species: Staphylococcus aureus
C>Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 20-Jun-2000
C:Accession: A37838
R:Yoshida, H.; Bogaki, M.; Nakamura, S.; Ubukata, K.; Konno, M.
J. Bacteriol. 172, 6942-6949, 1990
A:Title: Nucleotide sequence and characterization of the Staphylococcus aureus nora gene
A:Reference number: A37838; MUID:91072245; PMID:2174864
A:Accession: A37838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <YOS>
A:Cross-references: GB:D90119; NID:g216974; PIDN:BAAI4147.1; PID:g216975
C:Superfamily: tetracycline resistance protein

Query Match 7.9%; Score 174; DB 2; Length 388;
Best Local Similarity 20.5%; Pred. No. 1.1e-05;
Matches 85; Conservative 76; Mismatches 191; Indels 62; Gaps 15;

OY 5 NKQWLGRLHLIMGYIAIAVMTGDFELAFLSHYIKSLGTFPAEASFATLYGLAALS 64
Db 2 NKQIFVL-----YFNIFLIFLIGLIVLPVLPYKLDLGLTGSDDLGLVAAPALSQMI 54
OY 65 AMWGVAAEITTPQKTMIGFVLCVPHVLELVFGLGANGLLLEFGIGLAPFLY 124
Db 55 SPFGGLADKLKGLKILICIGLILFSVEFMFV---GHNSVLMLSRVIGMSAGMMP 110
OY 125 SFVIVIIHNVSESSSALGWMAVYSGIGVAGSYIPSTPIPMGEKGLMLALCFEA 184
Db 111 GVTGLINDISPHQKAKNFYMSAIIINSGL-ILGPGIGFMAEYSHNR-----PTFA 162
OY 185 G--GVIAIIS---LRH---VKTGCHMHLTPREKFAELSRVATLLYTNRIFLSSIV-RI 235
Db 163 GALDILAFIMISIVLIHPKPKSTSGFQKLEPO-----LLTKIMWKVFIPVITL 212
OY 236 INTLSLGEFAVIMPMFVDELGFTTSE-WLQVMAAFETFTTFSNIFMGIVAEKGMWVI 294
Db 213 VLSGLSFEFLYSLYIADKVNYPKDIATITGGIGFALQIYFED-----KFM 263
OY 295 RMFGCLGAASSLAF---YVMPQYFGHNYMAMIPALGTFVAFVMAAV---FPAL 348
Db 264 KYFSELFIAMSLYSVVLILVAFANGWSIML--ISFVFIFGDMIRAPITNYFSIA 321
OY 349 PKHGAISVYNLSAGNSNFIAPAIVALL-----PMESTIGVIAVITAYL 396
Db 322 GERGFAGGINSFTPTSMGNFGLPLAGALFDVHIEAPYMAIGVSLAGVIVLI 375

RESULT 8
G70042
multidrug-efflux transporter homolog yvka - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G70042
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Bertea
```

```
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Exlington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koelter, P.; Koningsstein, G.; Krogh, S.; Kunito, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scnl
A:Authors: Schleich, S.; Schreier, R.; Scifone, F.; Seliguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, E.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G70042
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-444 <KUN>
A:Cross-references: GB:299122; GB:AL009126; NID:g2636029; PIDN:CABI5538.1; PID:g26360
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvka
C:Superfamily: fosmidmycin resistance protein

Query Match 7.8%; Score 173.5; DB 2; Length 444;
Best Local Similarity 20.1%; Pred. No. 1.4e-05;
Matches 91; Conservative 87; Mismatches 185; Indels 89; Gaps 19;

OY 4 NKQWLGRLHLIMGYIAIAVMTGDFELAFLSH-YIKSLGTFPAEASFATLYGLA 62
Db 3 SRKEM-ALYISLALGAILVPINSTIMAVALSISHTNESI---ASITVAVVYLYVMA 57
OY 63 LSAMWGVAAEITTPQKTMIGFVLCVPHVLELVFGLGQA---NYGLILEVIGRLAY 119
Db 58 VTQYIAGKIDMT-GNKTM---YLMGV--GLPIASLGCALSPSLILLYFRALQAVG 110
OY 120 PLFLYSFVLIHNVSESSSALGWMAVYSGIGVAGSYIPSTPIPMGEKGLMLA 178
Db 111 ALTPNSIAIRHVVSERKLPKVFPGFG--LGAGLGAALGPFISILIDFSWHSIRVWV 168
OY 179 LAF-----CFAGVIAIISLRVIRP-GH----- 201
Db 169 IPIALALFTALMPQYKENSADPIDIGSLILAGSIVSILLTNKAPWGTIVSVL 228
OY 202 ---MHNLTPEKPAE-----LSRAVTLTYTNRIFLSSIVRIITLSLGEFAVIMPM 250
Db 229 ILFLVPLFPRREKTPRIIDFALFKST--FTNANLSV-----LSNLMKAVVAILMPL 281
OY 251 MEVDLGFTTSEWLVMAAFETFTTFSNIFMGIVAEKGMWVIIRMFEGCLMAASSLA 310
Db 282 FMTQFGLNLSNSGMAISVFISFMSASNMVGAOLHHRWGAKKIT-FLSFAMMAGANLFL 340
OY 311 YMPQYFGHNYMAMIRPAL-ALGTFVAFVMAAVFPALPEKHKGAISVYNLSAGNSN 368
Db 341 LLSS--SHSVLEFLMLSLILGLSAGVGLTSMQVSLATVPDGMGASVAGIFSTFRYFGST 398
OY 369 LAPAIVALLPMESTIGVIAVITAYLALVLL 400
Db 399 ISSAL-----IDLISGYHFLFMILFAV 420

RESULT 9
E71614
membrane transporter PFB0465C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71614
R:Gardner, M.J.; Tetteijn, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pereira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9604551
```


Db 226 -----NBAEPFLTLCVFIYLFSTFGANWPIGFLPTYLAGE-GFDGCVSNLMTAAFG 278
Qy 274 TIFSNIFWGIYAERKGMWRVIRWFGCLGMAASLAFYMPQFGHNWMM-----AMIPAL 330
Db 279 TIVGINWGLCADRIKLFKTF-----SIGLMSFLFIPLFRIPDONTILLGACLEGLMAT 334
Qy 331 GTFPAAYVP-MAAVFPALPEKHKAISVYNLSA-----GMSNFLAPAI 375
Db 335 NVGVCGLVPKFLYDFP-LEVRGLGTGL-IVNLATSGTENSMAATWIGLTMGLAALT 392
Qy 376 VLLPWFSTIGVVI 388
Db 393 IVAFWTATILLI 405

RESULT 12

C70082
Multidrug-efflux transporter homolog yx1H - Bacillus subtilisC:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: C70082

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koether, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portereid
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Satale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeder, R.; Scollone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wippl, A.; Yamano, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: C70082
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-400 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15890.1; PID:el186363;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yx1H

Query Match 7.3%; Score 161.5; DB 2; Length 400;
Best Local Similarity 21.5%; Pred. No. 9,7e-05;
Matches 93; Conservative 70; Mismatches 158; Indels 111; Gaps 18;

Qy 18 GYIAIVPMFGDGEFLAFLSHYIKSLGFTPAEASFATLGLAALASAWGVAAEII-T 76
Db 33 GYIVVPF-----SLYLEDLHFSGAIGIILSGSYVTQILRFPLGLSLDILFS 81
Qy 77 POKTMLI-GFVLMGVFHYLFVFLGLGOANYGLL--LFVIRGLAVPL-ELTSFIYI 130
Db 82 LKROLLAGGFEVSLSLLELMPD---SEFFVLARFPAITASTWMAITILA----- 132
Qy 131 IHNRENSSSALG--WYMAVYS-----VGIGVAGSYIPSTFIPIMGEMGLMALAFCE 183
Db 133 -HYNNNGNASKAMKIMOFFYVMPFASIVFCGLAAH-----LGRQVFFMALLASA 183
Qy 184 AGVYIAMISLRHVTPGHMNLPRKFAELSRVATL-LYTNRNIFLSIIRIINTLSLF 242
Db 184 AGVITCGFIODPSAPRANRGIRVNOYIKDRLPKLKLTF-----ILSMTHAALFTIVF 239
Qy 243 GFAVIMPMFVDELGFTTSEMLOVMAAFPTTFFSNIFWGIYAERKGMWRVIRWFGCLG 302
Db 240 GFT-----PLYMNOGLMGDAELLMWSAF-----LPH 267
Qy 303 AASLAFYVMPQYFGHNWMMAMIPALGLTFVAAPVMAA-----VF 344
Db 268 AAATLSFVPL-RTFSRIAYSMILISFAVTVGCLLVFPLAALFTVCITTHACIGLALGVF 326

Qy 345 PAL-----EPKHKGAISVYNLSAGMSNFLAPAIYVLLPWFSTIGVIATATYLL 396
Db 327 PLLSHVEISSARLKKMSVGFQSFYALGIFGLPRLLAGKIAQLIGLAGVFYAGSLAFA 386
Qy 397 AFVLCATIRVQ 408
Db 387 AF-----FVMLAQ 394

RESULT 13

AD0675

membrane transport protein STY1519 [imported] - Salmonella enterica subsp. enterica s

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AD0675

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: AD0675

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD01773.1; PID:g16502620; GSPDB:GN00176

C:Genetics:
A:Gene: STY1519

Query Match 7.2%; Score 159.5; DB 2; Length 428;
Best Local Similarity 20.3%; Pred. No. 0.00015;
Matches 99; Conservative 73; Mismatches 173; Indels 143; Gaps 19;

Qy 1 MSRRNKQMLGLPILMIGYIAIVPMFG--DGEFLAFLSHYI-KSLGFTPAEASFATLY 57
Db 1 MKRRTGMGLVY-----FLFIYMLNMYDRSALSTAPATELELGFNAEMGMITSAF 53
Qy 58 GLAALASAWGVAAEITTPQKTMILGFLVLMCVHVFVFLVGLGOANYGLLLEFGIRGL 117
Db 54 FIGALNLFIDGMA SDKVGPRTVFLAALLMSV-----FCGLTGL 93
Qy 118 AYPLELYSFIYV-----IHN-VSENSSSALGMYMAVYSVGAGSYI 161
Db 94 VTGLMTLIVLVLEGMAGPVSAGNKIINNWMISRKESATAIGFSAGSPVAGAVSG--- 150
Qy 162 PSFTIPIMGEMGLMALAPCFAGCVIAMISLRV-----KTPGHMNLPRKFAE 213
Db 151 ----PI--VGLLALSGLMPAPGILFLGLVWVLLMYFIVSDKPTMSKRLAEBRID- 201
Qy 214 LSRVATLLYTNRNIFLSIYRIINTLSLFGFAVIMPMFVDELGFTTSEMLOVMAAFET 273
Db 202 -----FENNEDVILSDDGATPSL---GYMMQPMVMAATTLAFFSYNI-----LFFFL 247
Qy 274 TIFS-----NIFWGIYAERKGMWRVIRW-FGCLGMAASLAFYVMPQYFGHNWMMAMIPAI 329
Db 248 TWFSYLNHSLDLDIKELISATVIPWVIGAGVGLVCSDVIYRIGN--ALLSRLI 304
Qy 330 LGTFVAAPVMAAVFPALPEKHKAISVYNLS----- 362
Db 305 LGVCLAGAANCVAASGVST--IGSAITLMSVSFLLYLTGPYIMAVIYDVHDKVGSV 362
Qy 363 -----AGMSNFLAPAIYVLLPWFSTIGVIATATYLLAFVLC-----AFIR 405
Db 363 GGAMHGLANTSGITGLPVT-----GFIYVQSGKYDAFYALAGAIATVSSLLVFP 412
Qy 406 VEDQGFSS 413
Db 413 VKSKGFKA 420

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 8.42384 seconds
(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNKQWGLPLHLIMGYI.....RVEQPGSSAPVTEKALNIS 424

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1917	86.7	427	1 RBT_KLEPN	052717 klebsiella
2	1907	86.3	425	1 DALY_KLEPN	052718 klebsiella
3	821	37.1	435	1 CSBX_BACSU	005390 bacillus su
4	709.5	32.1	414	1 YOAB_BACSU	034864 bacillus su
5	180.5	8.2	456	1 UHPT_CHLFR	084548 chlamydia t
6	174	7.9	388	1 NORA_STAM	P21191 staphylococ
7	158.5	7.2	455	1 UHPT_CHLFR	092703 chlamydia p
8	158.5	7.2	456	1 UHPT_CHLFR	092703 chlamydia m
9	158	7.1	413	1 MDCK_ACICA	P94131 acinetobact
10	151.5	6.9	428	1 YXIO_BACSU	P42306 bacillus su
11	144	6.5	458	1 TCR_STRAG	P13924 streptococc
12	143.5	6.5	595	1 BEHP_CORGL	P54582 corynebacte
13	143	6.5	454	1 YABR_ECOLI	P77726 escherichia
14	141	6.4	497	1 DTPY_LACLA	P36574 lactococcus
15	140	6.3	429	1 G6PU_HUMAN	043826 homo sapien
16	138	6.2	428	1 BCH2_RHOCA	P26171 rhodobacter
17	137.5	6.2	400	1 BMR2_BACSU	P39843 bacillus su
18	137	6.2	388	1 Y535_BUCAL	P57601 buchnera ap
19	137	6.2	405	1 PMRA_LACLA	P58120 lactococcus
20	136.5	6.2	405	1 YUHB_ECOLI	P39352 escherichia
21	136.5	6.2	532	1 COX1_RHOCA	P98059 rhodobacter
22	135	6.1	463	1 PGRP_SALTY	P12681 salmoneilla
23	133.5	6.0	395	1 NARK_BACSU	P46907 bacillus su
24	133	6.0	419	1 YHHS_ECOLI	P37621 escherichia
25	133	6.0	512	1 EMRB_ECOLI	P27304 escherichia
26	132.5	6.0	458	1 TCR_BACST	P07361 bacillus st
27	132.5	6.0	458	1 TCR_STAHY	P36890 staphylococ
28	132.5	6.0	458	1 TCR_STAHY	P11063 streptococ
29	132	5.9	430	1 RFBX_SALTY	P26400 salmoneilla
30	131.5	5.9	389	1 BMR1_BACSU	P33449 bacillus su
31	131.5	5.9	473	1 PHDK_NOCCK	024723 nocardioide
32	131.5	5.9	485	1 YJDL_ECOLI	P39276 escherichia
33	129.5	5.9	404	1 YYBF_BACSU	P37498 bacillus su

34	128.5	5.8	452	1 WZEY_SALTY	Q82396 salmoneilla
35	128.5	5.8	452	1 WZEY_SALTY	P37458 salmoneilla
36	128.5	5.8	459	1 TCR2_BACSU	P14512 bacillus su
37	127.5	5.8	433	1 TCR_STAHY	P02983 staphylococ
38	127	5.7	839	1 NAH3_DIDNA	Q28362 didelphis m
39	126.5	5.7	473	1 GLE_ZYMO	P21906 zymomonas m
40	126.5	5.7	515	1 VMT2_RAT	001827 rattus norv
41	126	5.7	422	1 EXUT_BACSU	034456 bacillus su
42	125.5	5.7	407	1 YB04_MAEIN	P71369 haemophilus
43	125	5.7	423	1 YDIN_ECOLI	P76198 escherichia
44	125	5.7	446	1 CITN_SALDU	P31603 salmoneilla
45	125	5.7	446	1 CITN_SALTY	P31604 salmoneilla

ALIGNMENTS

RESULT 1	ID	RTT_KLEPN	STANDARD:	PRT:	427 AA.
AC	052717				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Ribitol transporter.				
GN	RBT.				
OS	Klebsiella pneumoniae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Klebsiella.				
OX	NCBI_TaxID=573;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-1033-SP14 / KAY2026;				
RX	MEDLINE-97464425; PubMed-9324246;				
RA	Heuel H., Turgut S., Schmid K., Lengeler J.W.;				
RT	"Substrate recognition domains as revealed by active hybrids between				
RT	the D-arabinitol and ribitol transporters from Klebsiella				
RT	pneumoniae.";				
RL	J. Bacteriol. 179:6014-6019(1997).				
CC	- SUBCELLULAR LOCATION: Integral membrane protein (potential).				
CC	- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX				
CC	SUBFAMILY.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: AF045244; AAC26496.1;				
DR	InterPro: IPR004748; KEGGLU-permease.				
DR	InterPro: IPR003662; sub_transporter.				
DR	Pfam: PF00083; sugar_tr; 1.				
DR	TIGRFAMS: TIGR00897; 2A0118.1.				
KW	Transport; Sugar transport; Transmembrane.				
FT	DOMAIN 1 7				
FT	TRANSMEM 8 28				
FT	DOMAIN 29 51				
FT	TRANSMEM 52 72				
FT	DOMAIN 73 79				
FT	TRANSMEM 80 100				
FT	DOMAIN 101 107				
FT	TRANSMEM 108 128				
FT	DOMAIN 129 141				
FT	TRANSMEM 142 162				
FT	DOMAIN 163 171				
FT	TRANSMEM 172 192				
FT	DOMAIN 193 238				
FT	TRANSMEM 239 259				
FT	DOMAIN 260 263				
FT	TRANSMEM 264 284				

RESULT 3
CSBX_BACSU STANDARD: PRT: 435 AA.
AC 005390: 032056: 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Alpha-ketoglutarate permease.
GN CSBX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PY79.
RX MEDLINE=97177783; PubMed=9025289;
RA Gomez M., Cutting S.M.;
RT "BoC encodes a putative forespore regulator of the Bacillus subtilis
sigma K checkpoint."
RL Microbiology 143:157-170(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Tosato V., Bolotin A., Bertani I., Valentino I., Bruschi C.V.;
RT "A 17.8 kb segment in the spoV-nadC region of the Bacillus subtilis
168 chromosome: sequencing and ruv operon identification."
RL Submitted (DEC-1997) to the EMBL/Genbank/DBD databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartier M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier R., Boutsier L., Brans A., Braun M., Briganel S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Erlington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Gusepti G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tashob V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
RA Viari A., Mamput R., Medler E., Medler H., Weizenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Darchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RN [4]
RP CHARACTERIZATION OF EXPRESSION.
RX MEDLINE=97254445; PubMed=909855;
RA Gomez M., Cutting S.M.;
RT "Identification of a new sigmaB-controlled gene, csbX, in Bacillus
subtilis."
RL Gene 188:29-33(1997).
CC -i- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -i- DEVELOPMENTAL STAGE: EXPRESSED DURING THE STATIONARY PHASE OF CELL
CC GROWTH.
CC -i- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX

CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X93081; CA63620.1; -
CC DDB: Y15896; CAB75328.1; -
CC EMBL: 299118; CAB14736.1; -
CC Subtilist: BG11922; csbx.
CC InterPro: IPR004748; sub_transp.
CC InterPro: IPR003662; sub_transp.
CC Pfam: PF00083; sugar_tr_1.
CC TIGRfam: TIGR00897; 2A0118; 1.
CC Transport; Transmembrane; Complete proteome.
CC CYTOPLASMIC (POTENTIAL).
CC 1 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 2 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 3 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 4 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 5 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 6 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 7 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 8 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 9 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 10 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 11 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 12 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 0 -> E (IN REF. 1).
CC G -> A (IN REF. 1).
CC MISSING (IN REF. 1).
CC F -> S (IN REF. 1).
CC 364
CC SEQUENCE 435 AA; 47863 MW; 408A1A91E019417 CRC64;
SQ
Query Match 37.1%; Score 821; DB-1; Length 435;
Best Local Similarity 36.5%; Pred. No. 1, le-40;
Matches 146; Conservative 89; Mismatches 161; Indels 4; Gaps 2;
QY 9 LGPLHLIMGYIAIAVFMGTGDFELAFLSHYISLGFPAEASFATLVGLAALSAWS 68
DB 14 IGIPSHMWCYIGVIFMNGDGLGQMLSPFLVDHGSQMSLSLTMGIATISAMLS 73
QY 69 GVAVAEITTPKTMILGIFVLMCFHVLFVFGLCQANGYLILFYIRGLAVPLFLYSFY 128
DB 74 GTFVQTMGRPKMTYGLAFILGSAFICGAIWAIPIHMYPALGYSVALRGIDYPLFAVSFLV 133
QY 129 VIHNVRSENSSALGMYNAVYSGVGAGSYIPSPFIPIPMGEMGILMALACFPAGVI 188
DB 134 WVSYSSTSONILGRAVGMFMFMFCGLNVLPFYSVAVPAFGINILMSALLPVAAGIL 193
QY 189 AMISLRHVKTGGMHNLTPREKFAELSAVTLTYTNINIFLSSIVRIINLSLGFVIM 248
DB 194 ALFFNKDKFTPIOKD---QPKKELSKATIMFENPKVIGGVYKTIKINAGCFRAIFL 250
QY 249 PMKVEDELGFTTSEMIQVMAAFETTFISNIFGVIKAKGMKRVIRWEGCLGMAASSLA 308
DB 251 P-TYLARVGSVSEMIQIMGTLEFVNIVFNIIGAVGDKLGMNVTVMFEGVCGIFTLA 309

OY 309 EYYPQYFGHNWMMAMIPALGTFVAAFPMAFPALBPBKGAISVYNLSAGNSNF 368
DB 310 LLYTPQILGHQYWMVIMTACCYGAALGYPLSLPTLAPDNKGAAMSVNLGFGICAF 369
OY 369 LAPAIANVLLPWFSTIGVIAATATLALAFVLCAFIRVEQ 408
DB 370 IAPGVSLFTPLGAGGVIMFALYFSAFLTRFTISE 409

RESULT 4
YOAB_BACSU
ID YOAB_BACSU STANDARD: PRT; 414 AA.
AC 034864;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Putative transporter yoab.
GN YOAB
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
the tcrC and ochaB loci cloned in a yeast artificial chromosome";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Azevedo V., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Dainton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertlen R.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
RA Gilm S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoli E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosco V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viart A., Wambutt R., Wedler E., Wedler H., Welteregger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis".
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: AF027868; AAB84444.1;
CC DR EMBL: 299114; CAB13747.1;
CC DR EMBL: BG13473; yoab.
CC DR Subtilisin: IPR004748; Keloglu-permease.
CC DR InterPro: IPR003662; sub_transporter.
CC DR Pfam: PF00083; sugar_tr; 1;
CC DR TIGRPFAMs: TIGR00897; 2A0118; 1.
CC KM Hypothetical protein: Transport; Transmembrane; Complete proteome.
CC FT TRANSNM 1 12
CC FT TRANSNM 12 32
CC FT TRANSNM 32 47
CC FT TRANSNM 47 68
CC FT TRANSNM 68 75
CC FT TRANSNM 75 96
CC FT TRANSNM 96 107
CC FT TRANSNM 107 128
CC FT TRANSNM 128 136
CC FT TRANSNM 136 157
CC FT TRANSNM 157 167
CC FT TRANSNM 167 188
CC FT TRANSNM 188 219
CC FT TRANSNM 219 240
CC FT TRANSNM 240 255
CC FT TRANSNM 255 276
CC FT TRANSNM 276 287
CC FT TRANSNM 287 307
CC FT TRANSNM 307 316
CC FT TRANSNM 316 337
CC FT TRANSNM 337 343
CC FT TRANSNM 343 364
CC FT TRANSNM 364 375
CC FT TRANSNM 375 398
CC FT TRANSNM 398 414
CC FT TRANSNM 414 AA: 45376 MW: 4FAELICBDB744C4 CRG64;
CC SQ SEQUENCE

Query Match 32.1%; Score 709.5; DB 1; Length 414;
Best Local Similarity 33.6%; Pred. No. 2.5e-34;
Matches 139; Conservative 88; Mismatches 172; Indels 15; Gaps 4;

OY 9 LGPRLHIMGYIAIVMTGDFGLAFSLHYISLGTTPAEPAPFLYGLAALSAWVS 68
DB 5 IGIPRLAKGFLGVLEFMDGDEGLSLPFLIENGLTVQGSISPIYIALAIAWVS 64
OY 69 GYVAEITTPKMTLIGFLVLCYHNVLFLVGLQANGLLFLYGLAAYPLFLYSFIV 128
DB 65 GYVLEAFGARMTGMLLEFVIGTAFTVGFEBQLNPVYVYVYVGLGYPFAVSFLT 124
OY 129 YIHNVSSENSSLGMYAVSVIGVAGSYIPSPITPIMGEGTLMALACFPAGVI 188
DB 125 WVIYRTPQSKLSTAVGFWFVAYCIGMFVGAWSVYIKAFGLNLTMSIFWVCGAFT 184
OY 189 AMISLRHVKPRGHNNHNTPREKA-ELSAVNTLLYTRNFSLIYIINTLSLFGAVI 247
DB 185 ALF-----INDDREKKRRKRSFAELKGVTLITFPRLVLTGIRIINISIGTPRV 240
OY 248 MPWKEVDELGFTTSEWLVAAAFPTTIFSNIFMGVIAERKGMWRVIRWEGCLGMAASL 307
DB 241 LPMHMQH-GISINVMVLINGTIFLGNIVNLFGLVIGDRFGKNTVIMEGVGCGIFTV 299
OY 308 AFYYPQYFGHNWMMAMIPALGTFVAAFPMAFPALBPBKGAISVYNLSAGNSNF 367
DB 300 LLYTPQILGHQYWMVIMTACCYGAALGYPLSLPTLAPDNKGAAMSVNLGFGICAF 359
OY 368 LAPAIANVLLPWFSTIGVIAATATLALAFVLCAFIRVEQ 421
DB 360 FVGPALAMFTIGLVGAAGVWIMFALYFSAFLTRFTISE-----EKAV 404

RESULT 5
UHPT_CHLTR
ID UHPT_CHLTR STANDARD: PRT; 456 AA.

```

AC 084548;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexose phosphate transport protein.
GN CT5344.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Katman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sb.ch).
CC -----
DR EMBL: AE001325; AAC68146.1;
DR InterPro: IPR000849; G1PT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRFAMs: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; GLPT; 1.
KM Transport; Sugar transport; Transmembrane; Complete proteome.
FT TRANSMEM 34 54
FT TRANSMEM 70 90
FT TRANSMEM 113 133
FT TRANSMEM 161 181
FT TRANSMEM 185 205
FT TRANSMEM 257 277
FT TRANSMEM 302 322
FT TRANSMEM 331 351
FT TRANSMEM 363 383
FT TRANSMEM 394 414
FT TRANSMEM 421 441
FT TRANSMEM 456 AA; 51729 MW; E443610D65B7F8F CRC64;
SQ SEQUENCE

Query Match 8.2%; Score 180.5; DB 1; Length 456;
Best Local Similarity 22.3%; Pred. No. 0.00056;
Matches 105; Conservative 71; Mismatches 171; Indels 123; Gaps 25;

```

```

DB 240 LEEBARELSTKELFTYVLNKMWLFSPASFYIVRAVND-----WSALY 288
OY 272 -----FTTIFSN-----FMGIVAE-----KMGMYIRFGLGMAAS 306
DB 289 LIETKDYSTVKANLCVSLFEIGLFGMLAGWLSDTISKRCRPMNV-----FSLGLVSI 345
OY 307 LAEYVPOYGHNVMMAMIPALGFEVAFVMAAV-PPALPCKHGAISVYNLSAGM 365
DB 346 LGLMGTRDRF--VWMDGTPLFIIGFLL--FGQMMIGLAAELSHKAA-----GRASGF 397
OY 366 SNEFLAPAAVLLPWFSTIGVVIATYAL-----YLAFLVLCAT 404
DB 398 TG-----WFAVFGAFAAGYPLGRVADQWGMHGFEVALLLACALI 435

RESULT 6
NORA_STAM
ID NORA_STAM STANDARD; PRT; 388 AA.
AC P21191;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Quinolone resistance protein nora.
GN NORA OR SAV0695 OR SA0650 OR MM0657.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315),
OS Staphylococcus aureus (strain MM2), and
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620, 1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino T., Ito T.,
RA Kanamori M., Matsunari H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ul Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Tekenchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Ikama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91072245; PubMed=2174864;
RA Yoshida H., Bogaki M., Nakamura S., Ubukata K., Konno M.;
RT "Nucleotide sequence and characterization of the Staphylococcus
RT aureus nora gene, which confers resistance to quinolones."
RL J. Bacteriol. 172:6942-6949(1990).
RN [4]
RP SEQUENCE OF 230-388 FROM N.A.
RC STRAIN=209P;
RX MEDLINE=91058531; PubMed=2173911;
RA Ohshta Y., Hiramatsu K., Yokota T.;
RT "A point mutation in nora gene is responsible for quinolone
RT resistance in Staphylococcus aureus."
RL Biochem. Biophys. Res. Commun. 172:1028-1034(1990).
CC -1- FUNCTION: INVOLVED IN QUINOLONE RESISTANCE. MAY CONSTITUTE A
CC MEMBRANE-ASSOCIATED ACTIVE EFFLUX PUMP OF HYDROPHILIC QUINOLONES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

```

CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003360; BAB56857.1; -
DR EMBL: AP003311; BAB41883.1; -
DR EMBL: AP004824; BAB94522.1; -
DR EMBL: D90119; BAB14147.1; -
DR EMBL: M62960; AAB59089.1; -
DR PIR: A37838; A37838.
DR InterPro: IPR004734; Drug_resist.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR00880; 2_A_01_02; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 353 375 POTENTIAL.
FT TRANSMEM 352 352 D -> A (NO QUINOLONE RESISTANCE).
FT TRANSMEM 352 352 M -> V (IN REF. 4).
FT TRANSMEM 362 362 A -> D (IN REF. 4).
FT TRANSMEM 362 362 A -> D (IN REF. 4).
SQ SEQUENCE 388 AA; 42265 MW; C36177EAEADAC57 CRC64;
Query Match 7.9%; Score 174; DB 1; Length 388;
Best Local Similarity 20.5%; Pred. No. 0.0012;
Matches 85; Conservative 76; Mismatches 191; Indels 62; Gaps 15;
OY 5 NKOMLGLPLHIMGIAIVMTDGGFELAFSLHSHKSLGTPRAASAFLLYGLAALS 64
DB 2 NKQLEFVL-----YENIFLFLGIGLVIPVLYLKDLGLTGLLVAFAALSOMIT 54
OY 65 AMVSGVVAELITPQKTMIGFVLCVPHVFLVFGIGQANGLILFEYIGRLAVPLFLY 124
DB 55 SPFGSLADKLGKLLIGLILFSLVSSEFMFAV---GHNSVILMSLVIGMSGMYMP 110
OY 125 SFIYVYIHNVSSENSSALGMYMAVYSGIVAGSYIPSETPITMGENGTLMLALAFCA 184
DB 111 GVTGLINDISPSHQKAKNFYMSAIIINSGF-ILSGIGGFMAEVSHRM-----PEYFA 162
OY 185 G-GVIMAMIS---LRH---VKTGPHMHLTPREKFAELSRVTLTYTRNIFLSIV-RI 235
DB 163 GALTGLAFIMISIVLIHDPKSTSGFOKLEPO-----LTKIMKVFITPVVILTL 212
OY 236 INTLSLGFVAIPMMEVDELGFTTSE-WLVMAAFPTTIFSNIFMGIVAEKMGMMYVI 294
DB 213 VLSGLSAFETLYSLYADKNYSKPDISAIITGGIGFALFOIYFDP-----KFM 263
OY 295 RMFSCCLGMAASSLAF---YVMPQYFGHNYMMAMIPIALGTFVAAPVMAV---FPAL 348
DB 264 KYFSELTFFIAMSLLYSVVLLILFVANGYMSIML--ISFVEIGEDMIRAITNYESFIA 321
OY 349 PKHGAISIVNLSAGSNFLAPAIAYLL-----PMFSTIGVVIATYALYL 396
DB 322 GERGGFAGGLNSTFTSMGNFGLIAGALFDVHIETAPYMAIGVSLAGVIVLI 375
RESULT 7

UHP1_CHLPN STANDARD; PRT; 455 AA.
ID UHP1_CHLPN
AC Q927N9; Q9J019;
DT 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DN Probable hexose phosphate transport protein.
GN CPN0665 OR CPN082.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
ON NCPI_TaxID=83558;
RX MEDLINE-99206606; PubMed-10192388;
RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE-20150255; PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SUC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001649; AAD18804.1; -
DR EMBL: AE002171; AAF37968.1; -
DR EMBL: AP002547; BAA98872.1; -
DR TIGR: CP0082;
DR InterPro: IPR000849; G1PT_transporter.
DR TIGRfams: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; G1PT; 1.
KW Transport; Sugar transport; Transmembrane; Complete proteome.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
SQ SEQUENCE 455 AA; 51659 MW; F6404E2F1F4CA076 CRC64;

Query Match 7.2%; Score 158.5; DB 1; Length 455;
 Best Local Similarity 20.7%; Pred. No. 0.01;
 Matches 100; Conservative 74; Mismatches 160; Indels 149; Gaps 26;

```

OY 1 MSRRNKQW-LGLPLHLIMGYIAIAVFMGDDGELAFSLHYISLGFTPAEAS-FAFTLYG 58
DB 23 YKKKKYKRIIRIFYSMFICY--IFYFTKRSTFA-MPTLLIDLQGLIGSTLY- 78
OY 59 LAAALSAWVGVVAELITPQKTMILGFVLMCFVHVLFLVFGQANYGILLFYGI---- 114
DB 79 FSYGSKFVSGVSDQSNRFRYMAIGLMTGLTNIF--FGM-SSSYVLFALMWGLNGWF 134
OY 115 RGLAIP-----LFYSFIYVLIHNVRSSENSSALGWYMAVYSGIGVAGSYIP--SFTIP 167
DB 135 QGWMGPPCARLLTHWY-----AKSERGTWMSWTSVSHNIGALLIYLNGFIID 182
OY 168 IMGEMTLMALAFCFAGVIAIMISLR-----HYKTPGHMHN--LTREKFA 212
DB 183 YSGWGMAMVPGILICIGMGLVLINRLRDPQSLGLPRLEKYKRDPHHAHEKKSASEGTE 242
OY 213 ELISRAV-----TLLYTRNIFLSSIVRIITLSLFGAVIMPMMFVDELGFTTSEMIQ 265
DB 243 ELIERELSTRELLFTYVLTNOMLMF-----LAAASFIYIVRAVND----- 283
OY 266 VWAAREFTT-----IFSNIFMGIVAEKM-----GMBRVIRMGCL 300
DB 284 -WSALFLIETKHYAAVKANFCVSLFEIGLFGMLVAGLSDISIGNKCPMNVLSLG-- 340
OY 301 GMAASSLAFYVPOYF--GHNYMAMIPALIGT--FYNAFY-----PMAAVPALPEPKHK 352
DB 341 -----LFLAILGMWFSRSHNOMWD-----GTLLFVIGFLYGPQMMIGLAAELSLK 388
OY 353 GAISIVYNLSAGMSNFLAPALVLLPSTIGVIATYAL-----YLLAFVYC 401
DB 389 RAA-----GTASGETG-----WFAYPGATFAGYPLGKVVYDWGMKGFFALLAC 432
OY 402 APT 404
DB 433 ASI 435

RESULT 8
UHPT_CHLMU STANDARD: PRT; 456 AA.
ID UHPT_CHLMU
AC 09P1J18:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexose phosphate transport protein.
GN TC0831.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / Nig9;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Utecherack T., Berry K., Bass S.,
  Linher K., Weidman J., Kourti H., Craven B., Bowman C., Dodson R.,
  Gwinn M., Nelson M., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
  Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
  pneumoniae AK39."
RT Nucleic Acids Res. 28:1397-1406(2000).
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
  SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL: AE002349; AAF39631.1;
CC TIGR: TC0831;
DR InterPro: IPR000849; GLPT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; SugarLcr_1.
DR TIGRFAMs: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; GLPT_1.
KW Transport; Sugar transport; Transmembrane; Complete proteome.
FT TRANSMEM 34
FT TRANSMEM 70
FT TRANSMEM 113
FT TRANSMEM 133
FT TRANSMEM 161
FT TRANSMEM 185
FT TRANSMEM 257
FT TRANSMEM 302
FT TRANSMEM 331
FT TRANSMEM 362
FT TRANSMEM 394
FT TRANSMEM 421
FT TRANSMEM 441
SQ SEQUENCE 456 AA; 51651 MW; 830258E33D2F25A1 CRC64;

```

Query Match 7.2%; Score 158.5; DB 1; Length 456;
 Best Local Similarity 21.4%; Pred. No. 0.01;
 Matches 101; Conservative 73; Mismatches 170; Indels 129; Gaps 25;

```

OY 1 MSRRNKQW-LGLPLHLIMGYIAIAVFMGDDGELAFSLHYISLGFTPAEAS-FAFTLYG 58
DB 23 YKKKKYKRIIRIFYSMFICY--IFYFTKRSTFA-MPTLLIDLQGLIGSTLY- 78
OY 59 LAAALSAWVGVVAELITPQKTMILGFVLMCFVHVLFLVFGQANYGILLFYGI---- 114
DB 79 IYGISKFVSGVSDQSNRFRYMAIGLMTGLTNIF--FGM-SSSYVLFALMWGLNGWF 134
OY 115 RGLAIP-----LFYSFIYVLIHNVRSSENSSALGWYMAVYSGIGVAGSYIP--IP 167
DB 135 QGWMGPPCARLLTHWY-----AKSERGTWMSWTSVSHNIGALLIYLNGFIID 182
OY 168 IMGEMTLMALAFCFAGVIAIMISLRHYKTPGHMHNLTREKFAELSAVTLTYTRNI 227
DB 183 YSGWGMAMVPGILICIGMGLVLINRLRDPQSLGLPAIEKFKKEDSHRHEETADI 239
OY 228 PLSTIVRIINT-----LSLFGFAVIMPMMFVDELGFTTSEMIQVWAAPF 271
DB 240 LDEARELSTRELLFTYVLSKMKLMFLSFASFIYIVRAVND-----WSALY 288
OY 272 -----FTTIFSNL-----FMCIVAE-----KGMNRYVMFGCLGMAAS 306
DB 289 LIETKNYSTVKANLCVSLFEIGLFGMLVAGLSDTISKGRGPMNVLSLG----- 340
OY 307 LAFYVPOYFEGHN--YMMAMIPALIGTFAVYMAAV--FPALPEPKKAISVYNS 362
DB 341 LFLSILGLMGTHDRSTIMWADGAPFLITGFL--FGQMMIGLAAELSKKAA-----GTA 394
OY 363 AGMSNFLAPALVLLPWFSTIGVIATYAL-----YLLAFVLCAP 404
DB 395 SGTG-----WFAYFGAATFAGYPLGKVAQDWGMHGFVALLACALI 435

RESULT 9
MUCK_ACICA STANDARD: PRT; 413 AA.
ID MUCK_ACICA
AC P94131;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C1s/cis-muconate transport protein.
GN MUCK.

```

```

OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter..
OC NCBI_TaxID=471;
OX 11
RP SEQUENCE FROM N.A.
RC STRAIN-BD413 / ADP1:
RX MEDLINE=97440147; PubMed=9294455;
RA Williams P.A., Shaw L.E.;
RT "muck", a gene in Acinetobacter calcoaceticus ADP1 (BD413), encodes the
RT ability to grow on exogenous cis,cis-muconate as the sole carbon
RT source."
RL J. Bacteriol. 179:5935-5942(1997).
CC -I- FUNCTION: PROBABLE UPTAKE OF MUCONATE.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -I- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR	EMBL; U87258; AAC27117.1; '-'	
DR	InterPro; IPR004746; Bnzte_transport.	
DR	InterPro; IPR003662; sub_transporter.	
DR	Pfam; PF00083; sugar_trf_1.	
DR	TIGRFAMs; TIGR00895; 2A0115_1.	
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.	
DR	PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.	
KW	transport; Transmembrane; Inner membrane.	
FT	DOMAIN	1
FT	TRANSMEM	17
FT	DOMAIN	37
FT	TRANSMEM	53
FT	TRANSMEM	54
FT	TRANSMEM	54
FT	DOMAIN	75
FT	TRANSMEM	85
FT	TRANSMEM	86
FT	DOMAIN	107
FT	TRANSMEM	112
FT	DOMAIN	113
FT	TRANSMEM	133
FT	DOMAIN	134
FT	TRANSMEM	145
FT	TRANSMEM	146
FT	DOMAIN	167
FT	TRANSMEM	172
FT	TRANSMEM	172
FT	TRANSMEM	192
FT	DOMAIN	193
FT	TRANSMEM	228
FT	TRANSMEM	229
FT	DOMAIN	249
FT	TRANSMEM	250
FT	TRANSMEM	266
FT	TRANSMEM	267
FT	DOMAIN	288
FT	TRANSMEM	294
FT	DOMAIN	294
FT	TRANSMEM	315
FT	TRANSMEM	318
FT	TRANSMEM	319
FT	DOMAIN	339
FT	DOMAIN	340
FT	TRANSMEM	352
FT	TRANSMEM	353
FT	DOMAIN	377
FT	TRANSMEM	382
FT	TRANSMEM	383
FT	DOMAIN	404
FT	TRANSMEM	413
SO	SEQUENCE	413 AA; 45245 MW; 2088CE31C4CC65 CAC64;

QY 104 -ANGLLIFEGIGLAVPELY-SFVIVIIHNVRSSSSALGAWVAVYSGVAGSV 160
Db 105 FIOGG-VLRFPAASGLG---SLYACHTMLAEVYPTKRYTTLGTLOAGTFTGYIA-TL 159
QY 161 IPSTPIPMGEMGTLLMLAEFCFGVGIAMISLRHVTp-----GHMNLTPREKFAEL-S 215
Db 160 LAGMLIDPHG-----WVYL-FVYAILIVLAAVLHFFVPEPAAWOOSRLAPSKOTELVKT 213
QY 216 RAVLLV---TNRNIF-LSIVRIINILSLGFAVIVPMHFEDELGTSTSEMLOVAAAF 271
Db 214 SAPFLIFQDKRNRMFLIMALTAGELOPGYGVYVNMMPVLSLESLGKKPKEMTAHYVGT 273
QY 272 FTTIFSNIFFMGVIAEKGMKWRVIMFEGCLGNA--ASSLAF-----YAPQTFGHNYMA 322
Db 274 TAMILIGKLAEFMDKLG-RFTYVAFBAICTAIFLPLIVEYNSPDNITLLVYIEGLYG- 321
QY 323 AMIPALIGFVAVFMMAAFEPALPEBKHGAAISYVNLISAGSMNLPAPAAVALLPMFS 382
Db 332 --IFYGNATVMTESPITALGTAI-----GGAVNVRGLAA---IAPTIGFLASGGS 380
QY 383 T-IGVVIATYALLAVVLCATVEQ 408
Db 381 IGLGVVWGAAYFICGVIAPELIERKQ 407

RESULT	10
YXIO_BACSU	
ID	YXIO_BACSU
P42306;	STANDARD;
	PRT; 428 AA.
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Hypothetical protein yxio.
GN	YXIO OR S3AR
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=1423;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168 / BGSC1A1;
RX	MEDLINE=97124196; PubMed=8969509;
RA	Yoshida K.-I., Shundo K., Sano H., Seki S., Fujimura M., Yanai N., Mwa Y., Fujita Y.;
RT	"Sequencing of a 65 kb region of the <i>Bacillus subtilis</i> genome containing the <i>lic</i> and <i>cel</i> loci, and creation of a 177 kb contig covering the <i>gnt-sacxy</i> region";
RL	Microbiology 142:3113-3123(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RX	MEDLINE=98044033; PubMed=9384377;
RA	Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G., Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S., Borris R., Bourster L., Brans A., Braun M., Brignell S.C., Bron S., Broutlet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M., Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Eutian K.D., Eyring K.J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Funna S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Hotzappel S., Hosono S., Hullio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blancland M., Klein C., Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.W., Levine A., Liu H., Mesuda S., Mauel C., Medigue C., Meade N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Pairo V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M., Presecan E., Puljic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Seliguchi J., Sekowska A., Seior S.J., Serrior P., Shin B.S., Soldo B.,

QY 181 FCFAGVIAISLRHV-----TPGHMNLPRKFAELSAVTLTYTNRIFFLSIV 233
 Db 225 FLIYVSLSFLFVYKHRIKRVDPEDVDPGLGKNI-----LFMIGVL 263
 QY 234 ---RIIMTSLFSGFAVIMPMFVDELGFTTSEMLOVNAAPFTTIFSNIF---GIVAE 286
 Db 264 CGGIIFCTVA--GFVSNVPPYMKVDHQLSTAE--IGSVIIFPGTMSVITFGYIGILVD 318
 QY 287 KMGMYIRMEGCGIAGMAASSLAFFYMPQFCHNYMAMIPAIAG--TFVAAPVPMAAVF 344
 Db 319 RRGFLVYLN-IGVTFVLSVSLTAFLE--TTSWFTIITVFGGLSTFTVTP-IVVS 374
 QY 345 PALEPKRGAISVYNNAGSNFLAPAAVY-----LLPMFSTIGV---VIA 389
 Db 375 SSLKQDAGAGMSLINTFTSLSE--GTGIAIVGGLSLPLDQLRP---MGVDQSTYL 428
 QY 390 YTAIXYL 396
 Db 429 YSNLLL 435

RESULT 12

BETP_CORGL STANDARD; PRT; 595 AA.
 AC P54582;

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycine betaine transporter betP.
 GN BETP OR CGL0892.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 NCBI_Taxid=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Peter H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: HIGH-AFFINITY UPTAKE OF GLYCINE BETAINE (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE BCCP (TC 2.33) FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X93514; CAA6371.1; -;
 DR EMBL: AP005276; BAB98285.1; -;
 DR InterPro: IPR000060; BCCP_transporter.
 DR Pfam: PF02028; BCCP_1.
 DR ProDom: PD010111; BCCP_transporter; 1.
 DR TIGRFAMs: TIGR00842; bccp; 1.
 DR PROSITE: PS01303; BCCP; 1.
 KM Transport: Transmembrane.
 FT TRANSMEM 60 80
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 139 159 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.

FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 397 417 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 FT TRANSMEM 490 510 POTENTIAL.
 FT TRANSMEM 521 541 POTENTIAL.
 SO SEQUENCE 595 AA; 64209 MW; 8354B41C834CEEE2 CRC64;

Query Match 6.5% Score 143.5; DB 1; Length 595;
 Best Local Similarity 20.5%; Pred. No. 0.091;
 Matches 105; Conservative 70; Mismatches 154; Indels 183; Gaps 26;

QY 20 IAIIVMTGDFELAPLSHYKSLGFTPAEASFAFTLYGLAALSAWVGVAETTPK 79
 Db 69 IVALTVWVGIGKDFSTNFSSALSAVDNLGMAFILFG---TVFVFPTVIAASKFG 123
 QY 80 TMLG-----FVLMCFHVFLVFGIGQANVYGLLFLYGRIGLAYPLFLSYLV- 129
 Db 124 TIRLGIDEAPEFTVSW---ISMFAAGM--GIGLMFY--GTTEPLTFYRNGVPG 172
 QY 130 -IHNVRSESSSALGW---MAVYSGIGVAGSY-----IPSTIPIMGKGTL 175
 Db 173 HDEHNVGVAAMSTFMFHTLHPMAIYAI-VGLAIYSTFRVGRKQLSSAVPLGEGKAE 231
 QY 176 -WL-----ALAF--CFAG-----GVIAISLRHV 196
 Db 232 GWLGKILIDILAIITVFGTCSLGLALQIAGLSAANIITEDPSDWTIVGIVSLTAFI 291
 QY 197 KTPGHMNLPRKFAELS---RAVITLLYTNRIFFLSYRII----- 236
 Db 292 ---FSAISGVKGIQYL-SNANVLAALLAIFEVVGPVYSIINLLPGS 336
 QY 237 -NTLSLFGFAVIMPMFVDELGFTTSEMLOVNAAPFTTIFS-NIPMGIVAEKMGMY 293
 Db 337 IGNLYSNFPQAGRTAMSADG---TAGEWLGSWTFIYMAWMISRPVGNFLAIRISGRS 393
 QY 294 IR-----WFGCLGMAASSLAFFYMPQFCHNYV----- 321
 Db 394 IREFPLGLLVLPAGVSTVWSIFG--GTAIYF---FQNGESIWGDGAEEQLGLLHAL 447
 QY 322 ---MAMIRALNGTFVAAFVPMAAVFPALPEPKH-----KGAISVYNN 361
 Db 448 PGQIMGIIIMILLGTFTTISADSASTVWGMSTGSHGLEANKVYTAAMGATATIGTLTL 507
 QY 362 SAGMSNFLAPAAVLL---PW-FSTIGVYIA 389
 Db 508 LSGGDMLNSLNQNTIYAANPFLFVIGLWFA 539

RESULT 13

YAJR_ECOLI

ID YAJR_ECOLI STANDARD; PRT; 454 AA.
 AC P77726;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transport protein yajR.
 GN YAJR OR B0427.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Rau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12.";
 Science 277:1453-1474(1997);
 RN [2]

RP SEQUENCE FROM N.A.
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
 RA Duncan M., Federpriel N., Hyman R., Kalman S., Komp C., Kurd O.,
 RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.,
 RA Submitted (JAN1997) to the EMBL/GenBank/DBD databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
 CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AE000149; AAC73530.1; ALT_INIT.
 CC DR EMBL: U82664; AAB40183.1; ALT_INIT.
 CC DR Ecogen: E013614; Y4J9.
 CC DR InterPro: IPR003662; sub_transporter.
 CC DR Pfam: PF00083; sugar_tr.1.
 CC KM Hypothetical protein; Transport; Transmembrane; Inner membrane;
 CC Complete proteome.
 CC FT TRANSMEM 15 35 POTENTIAL.
 CC FT TRANSMEM 48 68 POTENTIAL.
 CC FT TRANSMEM 85 105 POTENTIAL.
 CC FT TRANSMEM 138 158 POTENTIAL.
 CC FT TRANSMEM 166 186 POTENTIAL.
 CC FT TRANSMEM 217 237 POTENTIAL.
 CC FT TRANSMEM 253 273 POTENTIAL.
 CC FT TRANSMEM 280 300 POTENTIAL.
 CC FT TRANSMEM 307 327 POTENTIAL.
 CC FT TRANSMEM 341 361 POTENTIAL.
 CC FT TRANSMEM 364 384 POTENTIAL.
 CC SQ SEQUENCE 454 AA; 48810 MW; B35BD/BEET5DE901 CRC64;
 Query Match 6.5%; Score 143; DB 1; Length 454;
 Best Local Similarity 23.0%; Pred. No. 0.078;
 Matches 98; Conservative 73; Mismatches 203; Indels 52; Gaps 18;

RESULT 14
 ID DPTT_LACLA STANDARD; PRT; 497 AA.
 AC P36574; Q9CHM6;
 DT 01-JUN-1994 (Rel. 29) Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Di-/tripeptide transporter.
 GN DPTT OR LT0702.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ML3;
 RC MEDLINE=94209318; PubMed=8157671;
 RA Hagling A., Kunj E.R.S., Leenhouts K.J., Poolman B., Konings W.N.,
 RT "The di- and tripeptide transport protein of Lactococcus lactis. A
 RL new type of bacterial peptide transporter.";
 RL J. Biol. Chem. 269:11391-11399(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RC MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RL "The complete genome sequence of the lactic acid bacterium Lactococcus
 RL lactis ssp. lactis IL1403.";
 RL Genome Res. 11:731-753(2001).
 RN [3]
 RP TOPOLOGY.
 RX MEDLINE=97327493; PubMed=9184160;
 RA Hagling A., van de Velde J., Poolman B., Konings W.N.;
 RT "Membrane topology of the di- and tripeptide transport protein of
 RT Lactococcus lactis.";
 RL Biochemistry 36:6777-6785(1997).
 CC -1- FUNCTION: PROTON-DEPENDENT UPTAKE OF DI- OR TRI-PEPTIDES.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE PRT FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U05215; AAA20660.1; ALT_INIT.
 CC DR EMBL: AE006303; AAK04800.1; -.
 DR PIR: A53620; A53620;
 DR InterPro: IPR000109; PPR2.
 DR InterPro: IPR005279; Peph_symporter.
 DR Pfam: PF00854; PPR2; 1.
 DR TIGRPFAM: TIGR00924; yjdl_subl_fam, 1.
 DR PROSITE: PS01023; PPR2_1; 1.
 DR PROSITE: PS01023; PPR2_2; 1.
 KW Peptide transport; Transport; Transmembrane; Complete proteome.
 FT DOMAIN 1 36 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 37 55 PROBABLE.
 FT DOMAIN 56 64 EXTRACELLULAR (PROBABLE).
 FT TRANSMEM 65 83 PROBABLE.
 FT DOMAIN 84 92 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 93 111 PROBABLE.
 FT DOMAIN 112 115 EXTRACELLULAR (PROBABLE).
 FT TRANSMEM 116 134 PROBABLE.
 FT DOMAIN 135 154 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 155 173 PROBABLE.
 FT DOMAIN 174 181 EXTRACELLULAR (PROBABLE).
 FT TRANSMEM 182 200 PROBABLE.
 FT DOMAIN 201 224 CYTOPLASMIC (PROBABLE).

```
FT TRANSNM 225 243 PROBABLE.
FT DOMAIN 244 254 EXTRACELLULAR (PROBABLE).
FT TRANSNM 255 273 PROBABLE.
FT DOMAIN 274 293 CYTOPLASMIC (PROBABLE).
FT TRANSNM 294 312 PROBABLE.
FT DOMAIN 313 335 EXTRACELLULAR (PROBABLE).
FT TRANSNM 336 354 PROBABLE.
FT DOMAIN 355 372 CYTOPLASMIC (PROBABLE).
FT TRANSNM 373 391 PROBABLE.
FT DOMAIN 392 425 EXTRACELLULAR (PROBABLE).
FT TRANSNM 426 444 PROBABLE.
FT DOMAIN 445 497 CYTOPLASMIC (PROBABLE).
FT CONFLICT 105 105 I -> V (IN REF. 1).
FT CONFLICT 219 219 S -> A (IN REF. 1).
FT CONFLICT 222 222 R -> K (IN REF. 1).
FT CONFLICT 226 226 L -> I (IN REF. 1).
FT CONFLICT 233 233 V -> L (IN REF. 1).
FT CONFLICT 237 237 L -> I (IN REF. 1).
FT CONFLICT 243 243 L -> I (IN REF. 1).
FT CONFLICT 323 323 D -> N (IN REF. 1).
FT CONFLICT 329 329 I -> F (IN REF. 1).
FT CONFLICT 356 356 L -> I (IN REF. 1).
FT CONFLICT 362 362 E -> D (IN REF. 1).
FT CONFLICT 380 380 I -> A (IN REF. 1).
FT CONFLICT 458 458 L -> I (IN REF. 1).
FT CONFLICT 485 485 V -> I (IN REF. 1).
SQ SEQUENCE 497 AA; 54844 MW; 348AE0D27960610C CRC64;
```

Query Match 6.4%; Score 141; DB 1; Length 497;
Best Local Similarity 19.2%; Pred. No. 0.11;

Matches 93; Conservativity 79; Mismatches 163; Indels 150; Gaps 21;

```
OY 1 MSRRNKKMGLPLHLI-----WGYIAVPMGDFELAFLSHYKSLGFTP 47
DB 4 LNKTEKTFEFOQPROLLLFQTEFEMERFSYGGMAILVY-----YLVALTTADNAGLPLK 58
OY 48 AEASFATLGLAALASWYSGVAE-IITPOKMLIGFVLMCFEHY-LFVLEGLQANY 105
DB 59 AQAMATYITIGALVYSTIYGWVADRGLGASRTIFLGILITLGHIALTPFELSG--- 115
OY 106 GLILFFYRGVLAFLFLYSFIVIIHNVSSENSALGWYMA-----VYSGVI 154
DB 116 -----LFVALFLIITIGMLKPNISNMVGHLSKDDSRDGTGFIYVGI 160
OY 155 GVAGSYPTPTIPMGEMGLMLALAFCFAGGVITAMIS-----LRHY-----KTGGMHNL 205
DB 161 NM-GSLIAPLIYGVGGVNVHLGFLSLAIGMIFALFAVYGRRLRHPPEIGREPSNPMDS 219
OY 206 TPREFAEISRAVTLTYTNRIPL-----SSIVRIITLSLGFVIMPMFVDELGF 258
DB 220 KARNRNL-ITLTIYVIAIIGFELLYQASPNANFINNINVLSTIG--IVPIIIF-VMMF 275
OY 259 TTSE-----WLOVMAAFFETITFSNIFMGIVAEMKGMNRV-----IRMFGLGM 302
DB 276 TSKVESDERRRKLTAIYIPLFSAL--VFMAIEQSSTIIAVGESRSNLDPTWFG--- 328
OY 303 AASLAFYVMPYF-----GHIYWM 322
DB 329 ---TFHIDPSWQQLNPLEFVLLSPITFVRLMKLGEROPSTIVKGLGIMLGISYLI 384
OY 323 AMIPALIGTEVAA-----FVPMMAVFPALPKHGAIAISYVNSLGSNFLAPAI--AVV 376
DB 385 MTLPGLLNGISGRASALMLVLMFAVOMAGE-----LLVSPGLSVSTKLAPAFOSOM 437
OY 377 LLEWF 381
DB 438 MAMWF 442
```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose 6-phosphate translocase (Glucose 5-phosphate transporter)
DE (PRO0685)
GN G6PT1 OR G6PT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GSD-IB CYS-339.
RC TISSUE=Urinary bladder;
RX MEDLINE=9808917; PubMed=9428641;
RA Gerin I., Veiga-Da-Cunha M., Achouri Y., Collet J.-F.,
RA van Schaftingen E.;
RT "Sequence of a putative glucose 6-phosphate translocase, mutated in
RT glycogen storage disease type Ib.";
RL FEBS Lett. 419:235-238(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GSD-IB ARG-118.
RX MEDLINE=99072316; PubMed=9856496;
RA Ihara K., Kuromaru R., Hara T.;
RT "Genomic structure of the human glucose 6-phosphate translocase gene
RT and novel mutations in the gene of a Japanese patient with glycogen
RT storage disease type Ib.";
RL Hum. Genet. 103:493-496(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
RX MEDLINE=99150335; PubMed=10026167;
RA Hiratawa H., Pan C.-J., Lin B., Moses S.W., Chou J.Y.;
RT "Inactivation of the glucose 6-phosphate transporter causes glycogen
RT storage disease type Ib.";
RL J. Biol. Chem. 274:5532-5536(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal liver;
RA Zhang C., Yu Y., Zhang S., Ouyang S., Luo L., Wei H., Zhou G.,
RA Zhang Y., Liu M., He F.;
RT "Functional prediction of the coding sequences of 9 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Li Y., van de Werve G.;
RT "Four different transcripts of putative glucose-6-phosphate
RT translocase in human leukocytes.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99148010; PubMed=10023055;
RA Gerin I., Veiga-Da-Cunha M., Noel G., Van Schaftingen E.;
RT "Structure of the gene mutated in glycogen storage disease type Ib.";
RL Gene 227:189-195(1999).
RN [7]
RP SEQUENCE FROM N.A.
RX PubMed=10323254;
RA Janecek A.R., Bosshard N.U., Mayatepek E., Schulze A., Gitzelmann R.,
RA Burchell A., Bosshard C.R., Janssen B.;
RT "Molecular diagnosis of type Ic glycogen storage disease.";
RL Hum. Genet. 104:275-277(1999).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Colon, and Lung;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [9]
RP VARIANT GSD-IB ARG-118.
RX MEDLINE=98342107; PubMed=9675154;
RA Kure S., Suzuki Y., Matsubara Y., Sakamoto O., Shintaku H.,
RA Ieshiki G., Hoshida C., Izumi I., Sakura N., Narisawa K.;
RT "Molecular analysis of glycogen storage disease type Ib:"

RT identification of a prevalent mutation among Japanese patients and
RT assignment of a putative glucose-6-phosphate translocase gene to
chromosome 11.¹
RL Biochem. Biophys. Res. Commun. 248:426-431(1998).
RN (10)
RP VARIANTS GSD-IB ARG-118 AND VAL-235 DEL.
RX PubMed-10482875.
RA Hou D.-C., Kure S., Suzuki Y., Hasegawa Y., Hara Y., Inoue T.,
Kida Y., Matsubara Y., Narisawa K.
RT "Glycogen storage disease type Ib: structural and mutational analysis
of the microsomal glucose-6-phosphate transporter gene".
RL Am. J. Med. Genet. 86:253-257(1999).
RN [11]
RP VARIANT GSD-IB GLU-149.
RA Lam C.-W., Tong S.-F., Lam Y.-Y., Chan B.-Y., Ma C.-H., Lim P.-L.
RT "Identification of a novel missense mutation (G149E) in the glucose-6-
phosphate translocase gene in a Chinese family with glycogen storage
disease Ib".
RL Hum. Mutat. 13:507-507(1999).
CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
reticulum (potential).
CC -1 ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2;
are produced by alternative splicing.
CC -1 TISSUE SPECIFICITY: MOSTLY EXPRESSED IN LIVER AND KIDNEY.
CC -1 DISEASE: DEFECTS IN G6PT1 ARE THE CAUSE OF GLYCOGEN STORAGE
DISEASE IB (GSD-IB).
CC -1 SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -1 CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 128.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y15409; CAA75608.1; -.
DR EMBL: AF078163; AAC72916.1; -.
DR EMBL: AF097831; AAD19898.1; -.
DR EMBL: AF111852; AAF16691.1; ALT_FRAME.
DR EMBL: AF110819; AAF37735.1; -.
DR EMBL: AF110820; AAF37736.1; -.
DR EMBL: Y17864; CAA76898.1; -.
DR EMBL: AF116864; AAD13111.1; -.
DR EMBL: AF116862; AAD13111.1; JOINED.
DR EMBL: AF116863; AAD13111.1; JOINED.
DR EMBL: BC002400; AAH02400.1; -.
DR EMBL: BC003589; AAH03589.1; -.
DR EMBL: BC014663; AAH14663.1; -.
DR Genew: HGNC:4061; G6PT1.
DR MIM: 602671; -.
DR MIM: 232220; -.
DR InterPro: IPR000849; G1PT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PFO0083; sugar_tr.1.
DR TIGRFAMs: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; G1PT. 1.
KW Transmembrane; transport; Sugar transport; Endoplasmic reticulum;
Alternative splicing; Glycogen storage disease; Disease mutation;
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT VARSPLIC 328 328
K -> KVAAEFTLALHPLAEITGFTTEH (IN ISOFORM
2).

OY	17	WGTAIAVF--MTGDGELAFLSHYKS-----LGFTPAEASFAFTLYG	58
Db	6	YGVRVTFVSAMFG-GVSLYVFNRKRTSEFVPMSLYEEIPLDKDGLGFTSSQSAAY----	60
OY	59	LAAALSAWVSVAEIIITPQ-----KIMLIG-----FYLMC-----VENHVLFLVYCLGANT	105
Db	61	--AISKVSVLESDQMSARULFESSGLLVNLGVNIFFRWSVTYPFALMFNGLAD---	114
OY	106	GILILEYGRGLAVPLFLYSFTVIINHVRSENSSSALGMWAYSGIVAGSYPSFT	165
Db	115	-----GLCMP-----PCGKYLKRWFEPSPQPTMMALLSTSMNLAGLGCP-IL	155
OY	166	IPIMGE---MGTLMLALFCFAGGVAMI SLRHVKTPGHMHNLTP--REKFAELSRAY	218
Db	156	ATIIAOSYMSRSTLASCALCVVASPCLLLIHNEPADVGLRNLDMPBSGGKSKLKEES	215
OY	219	TLLTNNRIEFLSFIRIINTLSLGFPAVIMPMMEVDLGFTTSMLOVMAAFPTTIFSN	278
Db	216	TL---DELLISPLYMLIST-----GYLUV-----EGVKTCCTDMGOFF-----	250
OY	279	IFMGIVAER-----MGMMRVIRFMFGCG---MAASSLAIFYMPOFGHNX	320
Db	251	----LIQRGGSALVGSSYSYMALEVGLVGSIAAGTYLSDRAMAKAGLSNTGNPR--HGL	303
OY	321	WMAMIPALALGTFYAA-----FVPMAAF-----PALERPKNKG	353
Db	304	LLFMWAGTVMYLFPRVTVTSDSPKLMIYLGAVERGSSYGPIALFGVIANESAPRYLCG	363
OY	354	A AISYYNLISAGMSNFLAPAIVLVLLPMFSTIGVIATATALLYLAFVCA	402
Db	364	TSHAIVGMANYGELAG-----LP-FSTIAKHYSSTAFWAEEVICA	405

Search completed: March 13, 2003, 16:53:47
Job time : 11.4238 secs

2) -

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:43:37 ; Search time 27.0265 Seconds

(without alignments)
3232.531 Million cell updates/sec

Title: US-09-802-208B-5

Perfect score: 2211

Sequence: 1 MSRRNKQWLGIPHLHMGVI.....RVEQPFSSAPYTEKALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

SPREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2211	100.0	424	2	09F4L5
2	625	28.3	408	16	08ZK49
3	181	8.2	462	16	09HWG1
4	176	8.0	456	16	08ZC53
5	175	7.9	388	2	053459
6	174	7.9	388	2	003325
7	173.5	7.8	444	16	034502
8	168.5	7.6	457	5	096186
9	164.5	7.4	386	17	058955
10	162.5	7.3	426	16	08ZK32
11	162.5	7.3	426	16	08Z7M8
12	161.5	7.3	400	16	094376
13	161.5	7.3	411	16	08ZPR3
14	159.5	7.2	428	16	08ZP66
15	159.5	7.2	428	16	08Z716
16	158.5	7.2	387	16	09CJ77

17	155.5	7.0	419	16	09HWT2	09HWT2 pseudomonas
18	155	7.0	397	17	09A129	09A129 streptococ
19	155	7.0	412	17	09HRX5	09HRX5 halobacteri
20	154.5	7.0	380	17	P95885	P95885 sulfolobus
21	154	7.0	388	2	09ZNA9	09ZNA9 staphylococ
22	153	6.9	405	2	066181	066181 arthrobacte
23	153	6.9	405	16	08Z257	08Z257 salmonella
24	152.5	6.9	392	2	09L804	09L804 pseudomonas
25	152	6.9	366	16	006473	006473 bacillus su
26	151	6.8	413	16	09K788	09K788 bacillus su
27	151	6.8	419	16	08ZLE4	08ZLE4 salmonella
28	150.5	6.8	409	16	09RTL7	09RTL7 deinococcus
29	150	6.8	401	16	09ZCF5	09ZCF5 listeria in
30	150	6.8	437	16	09HXPS	09HXPS pseudomonas
31	149.5	6.8	425	16	099RC3	099RC3 staphylococ
32	149.5	6.8	567	16	09PR59	09PR59 deinococcus
33	148.5	6.7	402	16	0912B6	0912B6 pseudomonas
34	148	6.7	432	10	09FKV1	09FKV1 arabidopsis
35	148	6.7	905	2	093PM1	093PM1 paracoccus
36	147	6.6	391	17	0979S3	0979S3 thermoplas
37	147	6.6	401	16	08Y7M4	08Y7M4 listeria mo
38	147	6.6	431	2	09ZRF4	09ZRF4 campylobact
39	147	6.6	431	16	09PMS5	09PMS5 campylobact
40	146.5	6.6	436	16	09XOF0	09XOF0 thermotoga
41	146.5	6.6	524	16	098GA1	098GA1 rhizobium 1
42	146.5	6.6	577	16	091134	091134 streptomyces
43	146	6.6	387	17	08U4E5	08U4E5 pyrococcus
44	145.5	6.6	381	17	058638	058638 pyrococcus
45	145	6.6	391	2	08VU72	08VU72 lactococcus

ALIGNMENTS

RESULT 1	
09F4L5	PRELIMINARY; PRT; 424 AA.
AC 09F4L5:	
DT 01-MAR-2001 (TREMREL. 16, Created)	
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)	
DE 01-JUN-2002 (TREMREL. 21, Last annotation update)	
DE Rbltol transporter.	
GN RPTT.	
OS Escherichia coli.	
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC Escherichia	
OX NCBI_Taxid=562;	
OX [1]	
RN SEQUENCE FROM N.A.	
RP SPRAIN-C.	
RA LaFayette P.R., Parrott W.A.:	
RT "A non-antibiotic marker for amplification of plant transformation	
RT vectors in E. coli."	
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
DR EMBL: AY005817; AAG01885.1; -	
DR InterPro: IPR001064; CrystalLin.	
DR InterPro: IPR004748; ketogluc.permease.	
DR TIGR: TIGR00897; 2A0118.1	
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA: UNKNOWN.1.	
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA: UNKNOWN.1.	
SQ SEQUENCE 424 AA: 46877 MW: 83962 CRC64;	

Query Match	100.0%; Score 2211; DB 2; Length 424;
Best Local Similarity	100.0%; Pred. No. 4.4e-150;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MSRRNKQWLGIPHLHMGVIATAVFMTGDGFEALFISHYKSLGTFPAEASFAFTLYGLA 60	
DB 1 MSRRNKQWLGIPHLHMGVIATAVFMTGDGFEALFISHYKSLGTFPAEASFAFTLYGLA 60	
QY 61 AALSAMVSGVAELITPQKTMILGFLVWCYFHYLFLVFGIGQANXYGILLFYGRLAYP 120	
DB 61 AALSAMVSGVAELITPQKTMILGFLVWCYFHYLFLVFGIGQANXYGILLFYGRLAYP 120	

```

QY 121 LFLYSEIVYIIHNVRSESSSALGWAVYSGVIGVAGSYIPSTPIPMGEMGLMLALA 180
DB 121 LFLYSEIVYIIHNVRSESSSALGWAVYSGVIGVAGSYIPSTPIPMGEMGLMLALA 180
QY 181 FCFAGVYIAMI SLRHVTPGHNHLPREKFAELSRATVTLTYTNRNIFLSSIVNIITLS 240
DB 181 FCFAGVYIAMI SLRHVTPGHNHLPREKFAELSRATVTLTYTNRNIFLSSIVNIITLS 240
QY 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTTIFSNIEMGIVAEKMGMRVIRMGCL 300
DB 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTTIFSNIEMGIVAEKMGMRVIRMGCL 300
QY 301 GMAASSAFYVMPQYFGHNWMMAMIPALCTFYAALFVMAAVPALEPRKKAIAISYN 360
DB 301 GMAASSAFYVMPQYFGHNWMMAMIPALCTFYAALFVMAAVPALEPRKKAIAISYN 360
QY 361 LSAGMSNFLAPALAVLLPWFSTIGVYIAYATLYLAFVLCAPFRVDPGFSAPYTEKA 420
DB 361 LSAGMSNFLAPALAVLLPWFSTIGVYIAYATLYLAFVLCAPFRVDPGFSAPYTEKA 420
QY 421 LNIS 424
DB 421 LNIS 424

```

RESULT 2

```

082K49 ID 082K49 PRELIMINARY; PRT; 408 AA.
AC 082K49;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative permease.
GN STM4434.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SCS1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McCalland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Gewai N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
DR EMBL: AE008908; AAL23254.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 408 AA; 44676 MW; DCE20DAAE075121B CRC64;

```

Query Match 28.3%; Score 625; DB 16; Length 408;
 Best Local Similarity 33.3%; Pred. No. 5.3e-37;
 Matches 134; Conservative 84; Mismatches 169; Indels 16; Caps 8;

```

QY 12 PLHLIMGYIAIVPMODGFELAFSLHYIKSLGTPPAASAFPLTGLAALASWVSGV 71
DB 16 PKFTWVSLCTFLPMLGCIENSWSLWMLNMGFDQAHIGOIFAGYIVAAITTSWLSGVC 75
QY 72 AEITTPQKTMIGFVLCVHFVLFVGLGQANGLILFYGIGLAVPLFLYSFIYII 131
DB 76 VDVGPRKKWVTVGRIYVLLASVAFNLALPSHDGAILVTYMLRGVGPLVCYSFLVRL- 134
QY 132 HNVRSESSSALG--WYAVYSGVIGVAGSYIPSTPIPMGEMGLMLALAFCPAGVIA 189
DB 135 -TOLDNHQOIGISTLEFWVYVNLGFTTIGPVASLIPELGHINWMAWAGKALL--GVLF 192
QY 190 MISLRHVKTGCHMHNLTPREK--FAELSRATVTLTYTNRNIFLSSIVNIITLSLFGAVI 247

```

```

DB 193 MLVIER-----NEFLPRTTPVEKELSGASIMWERPRIGLAVITKINLGYGFEVYV 247
QY 248 MPMFVDELGFTTSEMLQVMAAFETTTIFSNIEMGIVAEKMGMRVIRMGCLMAASSL 307
DB 248 LPLFLDLCK-HTELEAMSINWITFISNQVENIIFGWMGDKIGFRTQIIGSILITLVA 306
QY 308 AEYMPQYFGHNWMMAMIPALCTFYAALFVMAAVPALEPRKKAIAISYNNLSAGKSN 367
DB 307 IVYVPMIMGHNYVAFMLACLMGAGLAGEFVPMPLVPMAPDKGAANSVNFSGSLGN 366
QY 368 FLAPALAVLLPWFSTIGVYIAYATLYLAFVLCAPFRVDPG 410
DB 367 FVGPAL--VSYLAGFTGVYITMAGLITLFGSILVQLKV--PG 406

```

RESULT 3

```

09HWG1 ID 09HWG1 PRELIMINARY; PRT; 462 AA.
AC 09HWG1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable MFS transporter.
GN PA4233.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_Taxid=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufigle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AE004840; AAC07621.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane; Complete proteome.
SO SEQUENCE 462 AA; 48818 MW; C2178DC7E2DB45F0 CRC64;

```

Query Match 8.2%; Score 181; DB 16; Length 462;
 Best Local Similarity 22.6%; Pred. No. 2.7e-05;
 Matches 100; Conservative 75; Mismatches 182; Indels 86; Caps 20;

```

QY 13 LHLIMGYIAIVPMODGFELAFSLHYIKSLGTPPAASAFPLTGLAALASWVSGV 71
DB 20 LSLVAFARMLGMEVY-----LPVLATYGGDLAGATPALGIAAYGLTQAILOIPGTI 74
QY 72 AEITTPQKTMIGFVLCVHFVLFVGLGQANGLILFYGIGLAVPLFLYSFIYII 128
DB 75 SDRIGRPVIVG-----LILFAAGALAAANDSTIMGVAGNVLGAGAGISAAM 124
QY 129 VIHNV--RENSSSSALGWAVYSGVIGVAGSYIPSTPI-----PIW-----GEMGLMLA 178
DB 125 ALLSDTLRQQRHTKAM-----AMIGMSIGV-----SPAVAVVLCBPVTLHFGHLGFLMFT 174
QY 179 LAFCPAGVYIAMI SLRHVTPGHNHLPREKFAELSRATVTLTYTNRNIF 228
DB 175 AGMALVG--LLTLFVFPDPDHTQHRSSVARQALLPLKRGDILR----- 219
QY 229 LSSIVRIINTLSLPGFVIMPMVDELGFTTSEMLQVMAAFETTTIFSNIEMGIVAEK 288
DB 220 LDAGILILHAILMASF--VALPLALVHGGGLPKRQDHMMVYLTALLVGFAGVPIITYAEK 278

```

```

QY 289 GMMVIRRFEGCLGMAASSLAFYYPXQEGHHYMAAIPALALGFEVAFPMAPPALE 348
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 RRMKRV---LTGAVATLLACELFELVFGHSLANLVYGTV---VFETAFNLEBSPLSV 331
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 349 PK-----HKGAIVYNYLNSAGMSNFALPALAVILLPM-FSTIGVIVATYLLAF- 401
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 SKVPSPAGKGTAKMGVY---TSQFLGALVGLILGMMFQHGSLMVFICGAVLALMLA 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 AFIRVEDPFSSA---PYTERAL 421
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 IAVTMREPPYVTSIRLPLAPAL 410
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4	082C53	PRELIMINARY;	PRT;	456 AA.
ID	082C53			
AC	082C53;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	putative transporter.			
GN	YPO3159.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Yersinia.			
OX	NCBI_Taxid=632;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CO-92 / BIOVAR ORIENTALIS;			
RX	MEDLINE=21470413; PubMed=11586360;			
RA	Parkhill J., Wren B.W., Thomson N.R., Tittball R.W., Holden M.T.G.,			
RA	Prentice M.B., Sebithia M., James K.D., Churcher C., Mungall K.L.,			
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,			
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,			
RA	Feltwell T., Hamlin N., Holtroyd S., Jagers K., Karlyshev A.V.,			
RA	Leithner S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,			
RA	Smeeths M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;			
RT	"Genome sequence of Yersinia pestis, the causative agent of plague,"			
RL	Nature 413:523-527(2001).			
DR	EMBL: AJ141155; CAC92404.1; -			
DR	InterPro: IPR003662; sub_transporter.			
DR	Pfam: PF00083; sugar_tr; I			
DR	Hypothetical protein; Complete proteome.			
QO	SEQUENCE 456 AA; 48790 MW; 74D65EF5495B4095 CRC64;			

[illegible]

OY 337 FVPMAAVEPALEEP-----HKGAIISVTINSLAGSMNPLAIPAVLLP-MFSTIGVIAY 390
Db 318 FNWMAIILPSLISESPAGKYGTGMATYS-----TSQITGAIVGSGLOGMIFLEGADMYE 373

OY 391 TALYLFLAEVLCAFI-RVEQPGESSA 414
Db 374 AAGAIIALMWFAVSVMQEPPYVS 398

```

RESULT 5
053459
ID 053459 PRELIMINARY: PRT: 388 AA.
AC 053459;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE NORa.
CS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX NG E.Y., Trucksis M., Hooper D.C.;
MEDLINE=94379790; PubMed=8092836;
RT "Quinolone resistance mediated by norA: physiologic characterization
RT and relationship to flgB, a quinolone resistance locus on the
RT Staphylococcus aureus chromosome.";
RL Antimicrob. Agents Chemother. 38:1345-1355(1994).
CC 1- SUBCELLULAR LOCATION: INTERFAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: S74031; AAB31949.1;
DR InterPro: IPR004734; Drug_resist.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PFO0083; sugal_tr_1
DR TIGRfams: TIGR00880; 2_A_01_02.1.
KW Transmembrane.
SEQUENCE 388 AA: 42323 MW; 66C3883817827E8A CRC64;

```

Query Match	7.9%	Score 175;	DB 2;	Length 388;
Best Local Similarity	20.5%;	Pred. No. 6.2e-05;		
Matches	85;	Conservative	77;	Mismatches 190; Indels 62; Gaps 15;

Oy	5	NKQWLGRLHLIMGYATAIAVEMTGGDFELAFLSHYIKSLGFTPAEASFAETLLYGLAALS	64
Db	2	NKQIFVL-----YFNIFLFLIGLIVIPVLPVLLKDLGSDGLGLVAFAALSQMIT	54
Oy	65	AMVSGVAEITTPQKTMILGFLMKCVFHVLLVFLGIGQANYGILLFLFGIRLAPRLFLX	124
Db	55	SPFGGTLADKLGKLLICGLILFSLVSEFMFAV----GNFSLVLMLSRVIGMSGGMVP	110
Oy	125	SEIVIIINHVRSENSSALGWYWAYSVGIGAGSYTPEFTPIRIGEMQTMILALAFCA	184
Db	111	GVYGLADIISP HQAKNGVMSALIIINSF-ILGCGIGFMALVEVHRM-----PFLYA	162
Oy	185	G--GVIAIMIS--LRH--VKTPGHMHLTPREKFAELSRVAVTLTYNBNIFLSSIV-RI	235
Db	163	GALGILAEFLMSIVLIDHPKSTTSQFOKLEPQ-----LTKIKMKVETITVILLT	212
Oy	236	INTLSLFGRAVINPMMFVDELGETTSE-WLQYMAAFPTTIFPSNIFMCGVAAKMGMRVI	294
Db	213	VLSFGLSAETLYSLYTDKVMYSPKDISIALTGGICFALFOIYFPD-----KFM	263
Oy	295	RMFGCLIGMAASSLAF--YUMPQYEGHNVMYMMALPALAGTVAAFVPMVAAY--FPALF	348
Db	264	KYFSLLFTFIAMLLXSVVVLILLVANDYWSIML--ISVVEIGDMIRPALITNFSNIA	321
Oy	349	PKHKGAISVYNLSAGMSNFLAPATAVULL-----PMESTIGVIAVATALYLL	396
Db	322	GERQGFAGGLNSTFTSMGNETIGRLLAGLFDVHIEAPRIYMAAGVSLAGVAVILI	375

```

RESULT 6
003325
ID 003325 PRELIMINARY; PRT: 388 AA.
AC 003325;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE FLUOROQUINOLONE resistance NORAI199 protein (ALLELE NORAI199) (Allele
NORAI199(WT)).
GN NORAI199.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP ALLELE NORAI 1199, SEQUENCE FROM N.A.
RC STRAIN-SA-1199B, AND SA-1199;
RX MEDLINE=93297926; PubMed=8517696;
RA Kaatz G.W., Seo S.M., Ruble C.A.;
RT "Efflux-mediated fluoroquinolone resistance in Staphylococcus
aureus."
RL Antimicrob. Agents Chemother. 37:1086-1094(1993).
CC -1- FUNCTION: INVOLVED IN FLUOROQUINOLONE RESISTANCE. MAY CONSTITUTE A
MEMBRANE-ASSOCIATED ACTIVE EFFLUX PUMP OF FLUROQUINOLONES AND
OTHER DRUGS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- INDUCTION: FUNCTION OF NORAI199 PROTEIN AS MULTIDRUG EFFLUX
TRANSPORTER IS INHIBITED BY RESERPIN.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC DRUG RESISTANCE TRANSLOCASE
FAMILY.
CC -1- SIMILARITY: TO MULTIDRUG RESISTANCE PROTEIN FROM BACILLUS SUBTILIS
(P33449).
CC -1- CAUTION: THE SEQUENCE SHOWN IS THE SAME FOR ALLELE NORAI199 AND
ALLELE NORAI199(WILD TYPE). STRAIN SA-1199 CONTAINING ALLELE
NORAI199(WT) IS FLUROQUINOLONE-SUSCEPTIBLE. INCREASED
TRANSCRIPTION, AND NOT GENE AMPLIFICATION, OF NORAI199 IS THE
BASIS FOR FLUROQUINOLONE RESISTANCE IN STRAIN SA-1199B.
DR EMBL: M80252; AAA16158.1; -
DR EMBL: M97169; AAA26658.1; -
DR InterPro: IPR004734; Drug-resist.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRFAMs: TIGR00880; 2.A.01.02; 1.
KW Transmembrane; transport; Antibiotic resistance.
SQ
SEQUENCE 388 AA; 42236 MW; 81351490f77BD63 CRC64;

Query Match 7.9%; Score 174; DB 2; Length 388;
Best Local Similarity 21.0%; Pred. No. 7.3e-05;
Matches 87; Conservative 75; Mismatches 189; Indels 64; Gaps 16;

```

```

OY 5 NKOWGLPLHLINCYIAVFMGDDGELAFSLHYIKSLGFTPAEASPAFTLYGLAALS 64
DB 2 NKOLIV-----YFNFLFLFLGIGLVIPVLPYKLDLGLTGSDLGILVAAPALSGMII 54
OY 65 AMVSGVAEITTPKOTKLIGFVLCVPHVLFVGLGOANGLLLYGIRGLAYPLELY 124
DB 55 SPEGRTIADLKGKLLIICIGIL--FVSSEFMFAIQ--NELIIMLSRVGSGMAGMMP 110
OY 125 SFIIVIIHNVRENSSSALGMYAIVSVGIGVAGSYIPSEFTPIPMGEMGLMLAFCFA 184
DB 111 GVTLGIADISPSHQKAKNFYMSAIIINSGF-ILGPGIGGFAEYSHM-----PYFA 162
OY 185 G--GVIMT-----SLRHVKTTPGHHMNLTPREKFAELSRVATLLYTRNIFLSSIV-R 234
DB 163 GALTILAFINSIVLIHPKRVSTNG-FOKLEPQ-----LLTKINMKVFITPVILT 211
OY 235 IINTLSFGFAVIMPMPFVDELGFTTSE-MQVNAFAFFTITINIMGIVAEKMGMRV 293
DB 212 LVLSFGLSAETLSLTATKAVNSPKDISAITGGIGALFQIYFFD-----KF 262
OY 294 IRMEGICMAASSLAF--YYMPQYFGHNYMAMIPALIGTFFVAALVPMVA--FPAL 347
DB 263 MKYFSELTFIAMSLLTYSVIVLVLIADGYWTIMV--ISFVFLIGFDKIRPAITNYFSNI 320

```

```

OY 348 EPKHGAISVYVNLGAGNSNFLAPAIIVLL-----PWESTIGVIATYALYLL 396
DB 321 AGDRGFAAGINSTFTSMGNFGLIAGALFDVHIAPYIAVIGVSLAGVIYLI 375
RESULT 7
034502 PRELIMINARY; PRT: 444 AA.
ID 034502
AC 034502;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE YVKA.
GN YVKA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Soldo B., Rivoita C., Reynolds S., Manuel C.,
RA Karameita D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berrero M.C., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V.J., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertitz C., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Gilm S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Halberst G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Joris B., Karameita D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Prescecan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takauchi V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zamsstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AF017113; AAC67266.1; -
DR EMBL: Z99122; CAB15538.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane; Complete proteome.
SQ
SEQUENCE 444 AA; 47797 MW; 6A27F4EFB836CE8E CRC64;

```


Db 1 MIAKEFPWYS---ETTRPQKNALFSAMLVGDFDFMLIFYMYLIKADIGLADMEGA 56
 Oy 101 -----GOANYGILLFYGRGLAYPLFLYSPFVWIIHNVRSSSSALQWYA 148
 Db 57 FLATAAFTRPGFGLFGLLADKFGK-----PLMMMSIVAYSVGGLSGLASGVIMLTJS 112
 Oy 149 VYSGIGVAGSY-----IPSPFIPIGEMGTLM 176
 Db 113 RFTVGMAGKACASTYAVESMPKHLKSKASAFVSGFGIGNIIAAYMPSFAE-AYGM 171
 Oy 177 LALACFAGGVAMTSLR-----HKYTGCHMNLTPREFALSLRA 217
 Db 172 RAAPFVGLPLVLLVYIRARAPEKSEWEBAKLSGLGKHSQASMSVSLSMKGF----- 225
 Oy 218 VLLTYNRIPLFSIYRIINTLSFG-----PAVIMPMVDELGFTSEMLQWMAFFPT 273
 Db 226 -----NRAPFLTLVCFVLFVFSIFGANNPFIIGLPTIYAGE-GEPTGYVSNLMPTAAAG 278
 Oy 274 TFSNIFWGIYAEKMGWMMVIRMFGLGMAASSLAFTYMPQYFGHYWM---AMIPALAL 330
 Db 279 TVLGINVWGLCADRIGLKRTF---SIGLMSFLFIPLFRIPOQNYLLGLACFLGLMNT 334
 Oy 331 GFVNAFVP--MAAVFPALPEPKHKAISVYNSA-----GMSNFLAPAIAV 375
 Db 335 NVGVGGLVRFLEYDFP-LEVHGLGTGL-ITNLAATSGTFNSMAATWLTITMGLAALTF 392
 Oy 376 VLLPMFSTIGVVI 388
 Db 393 IYAFWTATILLII 405

RESULT 12

P94376 PRELIMINARY: PRT: 400 AA.

AC P94376: 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein yx1h.
 GN yx1h.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BGSC 1A1;
 RA MEDLINE-97124196; PubMed=8969509;
 RA Yoshida K., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT *Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the 11c and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacry region.*;
 RL Microbiology 142:3113-3123(1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA MEDLINE-98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bartero M.G., Bessières P., Bolotin A., Borchert S.,
 RA Boriss R., Boursier L., Brans A., Braum M., Brignell S.C., Bron S.,
 RA Braillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Chol S.K., Codani J.J., Connerton I.F., Cummins N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppl G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate I.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A.J., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Toononi A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wamburt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT *The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis*;
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: D83026; BAA11739.1;
 DR EMBL: 299123; CAB15890.1;
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 400 AA: 43963 MW: 185A568E76E4D38 CRC64;

Query Match 7.3%; Score 161.5; DB 16; Length 400;
 Best Local Similarity 21.5%; Pred. No. 0.00059;
 Matches 93; Conservative 70; Mismatches 158; Indels 111; Gaps 18;

Oy 18 GYIAIVFMTGDFELAFISHYKSGFPPAEASFMTLYGLAALSAWVGVAEII-T 76
 Db 33 GYIYVVF-----SLYEDLHFSYGAIGITLGSYQILRPPLGLSDILS 81
 Oy 77 POKTMLI-GFVLMCVFHVFLVFLVGLQANYGLT--LFYGRGLAYPL-FLYSPFVVI 130
 Db 82 LRQQLIAGFGSVLSLSLFMDP-----SPFVLAARLFAIGTASWVATILYA----- 132
 Oy 131 IHNVRSESSSALG--WYNAVYS-----VGIVAGSYIFSFTPIPGEMGTMLALAFCE 183
 Db 133 -HYFNNGNSAKMGIMQFETVMPQFASIVFCGLAAH-----LGRVPEWMAALASA 183
 Oy 184 AGGVIMISLRHKPRGHHNLPRKFAELSAVLT-LYTNRIPLFSIVRIINTLSF 242
 Db 184 AGLVICCFIODPSAPPANNGTIRVNOYIKDTLPRKLTFT---ILSPTAAVLFITVF 239
 Oy 243 GFVAVPMFVDELGFTTSEMILQWMAAFFTTFTFSNIFWGIYAEKMGWMMVIRMFGLCGM 302
 Db 240 GFT---PLYMNLGMDDELWMSAF-----LPH 267
 Oy 303 AASSLAFTYMPQYFGHYMMAMIPALALGTVAAYPMAA-----VF 344
 Db 268 AATLSEFVFL-RFTSRIVASVMLISPAVGVCLVYFLAALFTVCTHACIGLALGFVF 326
 Oy 345 PAL-----EKKHGGAISVYNLSAGMSNFLAPAAVAVLPMFSTIGVIAVATLYLL 396
 Db 327 PLLSHVVEISSARLMSWNGFYSPYALGIFLPLAGKIAOLIGLAGVYGAGSLAPA 386
 Oy 397 AFVLCAFIRVEQ 408
 Db 387 AF-----FVMAIQ 394

RESULT 13

O82PR3 PRELIMINARY: PRT: 411 AA.

AC O82PR3: 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative MFS family transport protein.

RP SEQUENCE FROM N.A.
 RC STRAIN-CT18;
 RX MEDLINE-21534947; Pubmed-11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Morgan K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., Mouton S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmone*
 RT *enterica* serovar Typhim CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627270; CAD01773.1;
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Complete proteome.
 SQ SEQUENCE 428 AA; 46519 MW; 5DBFE2EBD6E4AF74 CRC64;

Query Match 7.2%; Score 159.5; DB 16; Length 428;
 Best Local Similarity 20.3%; Pred. No. 0.00087;
 Matches 99; Conservative 73; Mismatches 173; Indels 143; Gaps 19;

QY 1 MSRRNNKQWLGRLHLMGTYIAIVFMTG--DGEELAFLSHYI-KSLGTPPAEASFAFTLY 57
 DB 1 MKRRTMGWLV-----FLFLIVMUNYDRSALSTITAPLRIEKELGFNAAEGMIFSAF 53
 QY 58 GLAALASAWYGVVAEITFPQKTMIGFVLCVFNFLVFGLGQANYGLILFYGIRGL 117
 DB 54 FICYALFNFDGWSADKVGKTVFLAALLMSV-----FCGLTGL 93
 QY 118 AYPLFLXFTIV-----IIHN-VRSNSSSALGWMYAVYVSGIGVASYI 161
 DB 94 VTGLWMTMLIVRYLFGNAEGPVSNAAGNKIINMWSRKESATAIGIFSAGSPLGAVSG--- 150
 QY 162 PSETTIPMGEMGFLMLALFCFAGVYAMISLRHV-----KTRGHMNLTPREKRAE 213
 DB 151 -----PI---VGLLALSLGWRPAFGITFLGLVWVLLMYFIVSDKPTMSKRLAPERID- 201
 QY 214 LSRAYVLLTYNNRIFLSSIVRIINFLSLFGFAVIMPMFVDELGFTTSEMLOVMAAFPT 273
 DB 202 -----FENHEDVILSDGGRATPSL---GYMKQPMWVATTLLAFSSNYI---LFFPL 247
 QY 274 TTFPS---NIFWGIVAEKMGMAVIRN-EGCLGMAASSLAFYTMPQYFGHNYMMAMIPATA 329
 DB 248 TWFPSTLNSLHDIDKEISIAIVIPWIGAIMVIGVCSVDIYRITGN---ALLSRRLI 304
 QY 330 LGTFVAAEVPMAAVFPALPEKHKGAISVYNLS----- 362
 DB 305 LGVCLAGAAVAVAGGTST--IGSAITLMSVSLPLLYITGPYMAVIODVVHKDKVGSV 362
 QY 363 -----AGMSNFLAPALAVLLPWFSTIGVAVYATAYLLAFVLC-----AFTR 405
 DB 363 GGAMHGLANISGIIIGPLVY-----GFIYQFSKGYDAFYLAGAIAIVSSLLVTF 412
 QY 406 VEQPGFSS 413
 DB 413 VKSKGFKA 420

Search completed: March 13, 2003, 16:55:16
 Job time : 31.0265 secs

ettu